

Universidad de Granada

Programa de Doctorado en Ciencias de la Educación



**UNIVERSIDAD
DE GRANADA**

Tesis doctoral

**LA INFERENCIA ESTADÍSTICA EN BACHILLERATO:
ANÁLISIS DE LAS PRUEBAS DE ACCESO A LA
UNIVERSIDAD Y DE LOS CONOCIMIENTOS DE
FUTUROS PROFESORES**

María del Mar López Martín

Granada, 2020

Universidad de Granada
Programa de Doctorado en Ciencias de la Educación



**UNIVERSIDAD
DE GRANADA**

**LA INFERENCIA ESTADÍSTICA EN BACHILLERATO:
ANÁLISIS DE LAS PRUEBAS DE ACCESO A LA
UNIVERSIDAD Y DE LOS CONOCIMIENTOS DE FUTUROS
PROFESORES**

Memoria de Tesis Doctoral realizada bajo la dirección de las doctoras Carmen Batanero Bernabeu y María Carmen Díaz Batanero, que presenta Dña. María del Mar López Martín para optar al grado de Doctora en Ciencias de la Educación por la Universidad de Granada.

Fdo. María del Mar López Martín

Vº Bº de las Directoras:

Fdo. Carmen Batanero Bernabeu

Fdo. María Carmen Díaz Batanero

Editor: Universidad de Granada. Tesis Doctorales
Autor: María del Mar López Martín
ISBN: 978-84-1306-770-4
URI: <http://hdl.handle.net/10481/66672>

Reconocimiento

Investigación realizada en el marco del Grupo PAI, FQM-126 (Junta de Andalucía), el proyecto de investigación EDU2016-74848-P (AEI, FEDER), el proyecto de investigación PID2019-105601GB-100/AEI/10.13039/501100011033, el Departamento de Educación de la Universidad de Almería y el Departamento de Didáctica de la Matemática de la Universidad de Granada.

AGRADECIMIENTOS

En estas líneas me gustaría manifestar mi agradecimiento a todas las personas que han puesto su granito de arena en el proceso de elaboración de esta memoria de tesis.

Quisiera manifestar mi más profundo agradecimiento a mis directoras de tesis por su constancia, compromiso y dedicación en una etapa que no ha sido nada fácil para ellas. El asesoramiento realizado y el constante ánimo y confianza depositada en mí han sido elementos fundamentales para que hoy pueda decir, ¡objetivo conseguido!

Gracias a mis compañeros del Departamento de Educación de la Universidad de Almería y a los compañeros del Departamento de Didáctica de la Matemática de la Universidad de Granada. Y como no, mencionar también a mi gran familia del Departamento de Métodos Cuantitativos para la Economía y la Empresa de la Universidad de Granada donde con mi grandes amigos Lina, Román, Víctor y Teresa comencé mi etapa en la Universidad. También quiero mencionar a José García Pérez por haberme animado a adentrarme en el mundo universitario. He pasado por muchos departamentos, pero en todos ellos me he sentido siempre arropada y feliz. ¡Gracias!

Por último, y no menos importante, agradezco a mi familia todos los momentos de alegría, de tristeza y de éxitos que habéis compartido conmigo, estando siempre ahí para dar apoyo.

Dar las gracias a Pablo y a mi pequeña ÁFRICA por todos las horas que no he podido estar con vosotros, entre la plaza de titular, la situación de pandemia y el término de este trabajo he renunciado a ese tiempo. Tu sonrisa y tus palabras de ánimo, “Mamá ya te queda poco. Estoy muy orgullosa de ti”, me han dado fuerza para seguir para adelante. Os quiero.

RESUMEN

Esta tesis doctoral se centra en el análisis de los problemas de inferencia estadística propuestos en las pruebas de acceso a la universidad y en la evaluación de algunos componentes del conocimiento didáctico-matemático de los futuros profesores sobre inferencia estadística. Se utiliza como marco teórico el enfoque ontosemiótico de la cognición e instrucción matemática y del conocimiento didáctico-matemático del profesor, elaborado en el mismo marco teórico. También nos basamos en los documentos curriculares vigentes y las investigaciones previas sobre la comprensión de conceptos de inferencia. Se realizan dos estudios empíricos que parten de dichos fundamentos:

- En el Estudio 1 se clasifican los problemas propuestos en Andalucía en las pruebas de acceso a la universidad de Matemáticas Aplicadas a las Ciencias Sociales II a lo largo de 12 años. Para cada uno de los campos de problema identificados, se realiza el análisis semiótico de la resolución de uno de ellos, identificando las prácticas y configuraciones de objetos matemáticos relacionados. Asimismo, se ha analizado el modelo de distribución de la población de partida en el ejercicio; el parámetro que se pide estimar y el contexto en el que se presenta el problema. A partir del estudio realizado, se lleva a cabo una reflexión sobre algunos componentes de la idoneidad didáctica de las pruebas de acceso a la universidad.
- En el Estudio 2 se evalúan algunos elementos del conocimiento didáctico-matemático que poseen un grupo de futuros docentes de Educación Secundaria Obligatoria y Bachillerato sobre inferencia estadística. A partir de sus respuestas abiertas a un cuestionario se analizan los conocimientos matemáticos de los participantes respecto al contraste estadístico de hipótesis e intervalo de confianza, identificando y clasificando sus principales conflictos semióticos. Se proporciona también una síntesis de los conocimientos matemáticos de los futuros profesores sobre inferencia estadística. El estudio se completa con la evaluación de los conocimientos sobre los errores que pueden cometer los estudiantes, que serían parte de la componente cognitiva de su conocimiento didáctico-matemático. También se investiga la forma en que explicarían el intervalo de confianza a sus estudiantes, que nos da indicaciones sobre una parte de la componente mediacional de su conocimiento didáctico-matemático.

Las principales aportaciones del trabajo son: a) la caracterización de las variables que definen los problemas de inferencia propuestos en las pruebas de acceso a la universidad en el periodo analizado y de su distribución temporal; b) la información sobre diferentes componentes del conocimiento de los futuros profesores, sus conflictos cognitivos y sus carencias; c) la síntesis de las investigaciones previas. Otra contribución son las publicaciones derivadas de la tesis.

ABSTRACT

This doctoral thesis focuses on the analysis of the statistical inference problems proposed in the university entrance tests, and on the evaluation of some components of prospective teachers' didactic-mathematical knowledge on statistical inference. The ontosemiotic approach to mathematics cognition and instruction and the teacher's didactic-mathematical knowledge are used as theoretical framework. We also rely on the analysis of current curricular documents and on the synthesis of previous research on understanding inference concepts. Two empirical studies are carried out that start from these foundations:

- In Study 1, we analyse the problems proposed in Andalusia in the entrance to university tests of Applied Mathematics to Social Sciences II over 12 years. For each of the problem fields identified, we carry out a semiotic analysis of the resolution of one of these problems, by identifying the practices and configurations of related mathematical objects. Likewise, the distribution model of the population, the parameter to be estimated and the problem context are analysed along the years. Based on the analysis carried out, a reflection is carried out on some components of the didactic suitability of the university entrance tests.
- In Study 2 some elements of didactic-mathematical knowledge of a group of prospective teachers of compulsory secondary education and high school on statistical inference are evaluated. Based on their open responses to a questionnaire, the participants' mathematical knowledge regarding the statistical hypothesis test and confidence intervals is analysed, by identifying and classifying their main semiotic conflicts. A synthesis of prospective teachers' mathematical knowledge on statistical inference is also provided. The study is completed with the evaluation of the teachers' knowledge about their students' errors, which is part of the cognitive component of their didactic-mathematical knowledge. The way they would explain the confidence interval to their students is also analysed, which gives us indications about a part of the mediational component of their didactic-mathematical knowledge.

The main contributions of this thesis are: a) the characterization of the variables that define the inference problems proposed in the university entrance tests in the period analysed and their temporal distribution; b) information on different components of prospective teachers' knowledge, their cognitive conflicts and their shortcomings; c) the synthesis of previous research. Another contribution is the set of publications derived from the thesis.

Índice

INTRODUCCIÓN I

PLANTEAMIENTO DEL PROBLEMA 1

1.1. INTRODUCCIÓN.....	1
1.2. INFERENCIA ESTADÍSTICA Y SU IMPORTANCIA.....	2
1.2.1. Desarrollo de la inferencia y su papel en el avance de las ciencias	4
1.2.2. Importancia en la formación del estudiante	6
1.3. LA INFERENCIA ESTADÍSTICA EN EL CURRÍCULO DE BACHILLERATO	8
1.4. PRUEBAS DE EVALUACIÓN DEL BACHILLERATO PARA ACCESO A LA UNIVERSIDAD	11
1.4.1. Las pruebas de acceso a la universidad	13
1.5. MARCO TEÓRICO.....	17
1.5.1. Tipos de objetos matemáticos	17
1.5.2. Significados institucionales y personales y sus tipos	19
1.5.3. Comprensión y evaluación en el EOS.....	22
1.5.4. Función semiótica y conflicto semiótico	24
1.5.5. Idoneidad didáctica y sus componentes	25
1.5.6. Modelo del conocimiento del profesor en el EOS	27
1.6. OBJETIVOS E HIPÓTESIS DE LA INVESTIGACIÓN	31
1.6.1. Objetivos del trabajo	32
1.6.2. Hipótesis o expectativas iniciales	34
1.7. ORGANIZACIÓN DE LA INVESTIGACIÓN	35

ANTECEDENTES DE LA INVESTIGACIÓN 39

2.1. INTRODUCCIÓN.....	40
2.2. INVESTIGACIONES SOBRE COMPRESIÓN DEL CONTRASTE DE HIPÓTESIS	42
2.2.1. Críticas al uso del contraste de hipótesis	43
2.2.2. Comprensión del concepto de hipótesis	44
2.2.3. Comprensión del nivel de significación y el p -valor	45
2.2.4. Comprensión del proceso de contraste.....	48

2.3. COMPRESIÓN DEL INTERVALO DE CONFIANZA	49
2.3.1. Definición del intervalo y del nivel de confianza	49
2.3.2. Propiedades del intervalo de confianza.....	52
2.4. DIFICULTADES CON OTROS CONCEPTOS DE INFERENCIA	53
2.4.1. Comprensión de la distribución muestral	54
2.4.2. Comprensión del teorema central del límite	58
2.5. ANÁLISIS DE PROBLEMAS DE INFERENCIA ESTADÍSTICA.....	59
2.5.1. Análisis de la inferencia estadística en libros de texto.....	60
2.5.2. Análisis de las pruebas de acceso a la universidad	61
2.5.2.1. Análisis de las pruebas de matemáticas	62
2.5.2.2. Estudio de la inferencia estadística en las pruebas de acceso a la universidad	65
2.5.3. Análisis de otras pruebas de evaluación	66
2.6. CONCLUSIONES	67
2.6.1. Conflictos identificados en la literatura.....	68
2.6.2. Posibles causas de los conflictos descritos	69
2.6.3. Conclusiones sobre las investigaciones relacionadas con las pruebas de acceso a la universidad.....	71

LA INFERENCIA ESTADÍSTICA EN LAS PRUEBAS DE ACCESO A LA UNIVERSIDAD

.....73

3.1. INTRODUCCIÓN	74
3.2. OBJETIVOS DEL ESTUDIO 1	76
3.3. HIPÓTESIS O EXPECTATIVAS INICIALES	78
3.4. METODOLOGÍA.....	79
3.4.1. Descripción general de la metodología	79
3.4.2. Muestra de problemas analizados	81
3.4.3. Variables analizadas	82
3.5. CAMPOS DE PROBLEMAS, PRÁCTICAS MATEMÁTICAS Y CONFIGURACIÓN DE OBJETOS MATEMÁTICOS ASOCIADOS	85
3.5.1. Análisis semiótico de problemas propuestos	86
3.5.2. Síntesis del análisis de los campos de problemas.....	106
3.5.3. Estudio comparativo entre el significado pretendido y evaluado.....	112

3.6. MODELO PROBABILÍSTICO DE LA POBLACIÓN DE PARTIDA	113
3.7. PARÁMETRO POBLACIONAL OBJETO DE ESTUDIO	117
3.8. CONTEXTO DEL PROBLEMA	119
3.9. ANÁLISIS ESTADÍSTICO DE LOS PROBLEMAS	122
3.9.1. Campos de problemas.....	122
3.9.2. Modelo probabilístico de la población de partida.....	124
3.9.3. Parámetro poblacional objeto de estudio.....	126
3.9.4. Contexto del problema.....	127
3.9.4.1. Estudio comparativo entre el campo de problema y el contexto.....	129
3.10. CONCLUSIONES DEL ESTUDIO 1.....	131
3.10.1. Idoneidad didáctica de los problemas de inferencia estadística en las pruebas de acceso a la universidad	131
3.10.2. Conclusiones sobre los objetivos.....	136
3.10.2. Conclusiones sobre las hipótesis	137

EVALUACIÓN DE CONOCIMIENTOS DE FUTUROS PROFESORES 141

4.1. INTRODUCCIÓN.....	142
4.2. OBJETIVOS DEL ESTUDIO 2.....	143
4.3. HIPÓTESIS O EXPECTATIVAS INICIALES	146
4.4. METODOLOGÍA.....	148
4.4.1. Muestra y contexto	149
4.4.2. Cuestionario propuesto.....	152
4.4.3. Método de análisis de los datos.....	156
4.5. CONOCIMIENTO MATEMÁTICO SOBRE EL CONTRASTE DE HIPÓTESIS	157
4.5.1. Planteamiento de las hipótesis.....	157
4.5.2. Elección del tipo de contraste	164
4.5.3. Desarrollo del procedimiento	169
4.5.4. Determinación del p -valor.....	175
4.5.5. Interpretación de los resultados.....	178
4.5.6. Síntesis de conocimientos sobre el contraste de hipótesis.....	183
4.5.7. Conclusiones sobre el conocimiento común del contraste de hipótesis ..	185
4.6. CONOCIMIENTO MATEMÁTICO SOBRE EL INTERVALO DE CONFIANZA	188
4.6.1. Planteamiento del intervalo de confianza	189

4.6.2. Cálculo del intervalo	193
4.6.3. Interpretación del intervalo	197
4.6.4. Definición	200
4.6.5. Propiedades que se recuerdan	202
4.6.6. Síntesis de conocimientos sobre el intervalo de confianza	205
4.6.7. Conclusiones sobre el conocimiento común del intervalo de confianza...	208
4.7. CONOCIMIENTOS DIDÁCTICOS SOBRE INFERENCIA ESTADÍSTICA	209
4.7.1. Errores en el planteamiento de las hipótesis	211
4.7.2. Errores en la selección del contraste de hipótesis e intervalo de confianza	212
4.7.3. Errores conceptuales	214
4.7.4. Errores procedimentales	218
4.7.5. Errores de interpretación	222
4.7.6. Formas de explicar el intervalo de confianza en el aula	225
4.7.7. Síntesis de conocimientos didácticos	228
4.7.8. Conclusiones sobre el conocimiento didáctico sobre inferencia estadística	230
4.8. CONCLUSIONES DEL ESTUDIO 2	231
4.8.1. Conclusiones sobre los objetivos	232
4.8.2. Conclusiones sobre las hipótesis	235
CONCLUSIONES GENERALES	239
5.1. INTRODUCCIÓN	239
5.2. CONCLUSIONES RESPECTO A LOS OBJETIVOS DEL TRABAJO	240
5.3. CONCLUSIONES SOBRE LAS HIPÓTESIS	243
5.4. PRINCIPALES APORTACIONES	245
5.5. LIMITACIONES DEL ESTUDIO Y FUTURAS LÍNEAS DE INVESTIGACIÓN	246
5.6. IMPLICACIONES PARA LA ENSEÑANZA	247
REFERENCIAS.....	249
ANEXO. TRABAJOS QUE DESARROLLAN ESTA MEMORIA.....	264

INTRODUCCIÓN

Desde el punto de vista formal, la inferencia estadística o estadística inferencial es un conjunto de técnicas y métodos que permiten inducir el comportamiento de una determinada población a partir de la información aportada de una muestra representativa de la misma.

La inferencia estadística se creó con objeto de dar respuesta al problema consistente en obtener un conocimiento general (una nueva teoría científica) a partir del análisis de casos particulares (inducción empírica), esto es, de la búsqueda de métodos que permitan justificar este tipo de razonamiento inductivo y la extensión de sus conclusiones (Black, 1979; Mayo y Cox, 2006; Vallecillos, 1994). Este hecho ha provocado que inferencia estadística juegue un papel fundamental en el desarrollo de las ciencias empíricas, que no pueden justificar sus conclusiones mediante razonamiento deductivo.

Durante las últimas décadas son diversas las investigaciones que han puesto el foco de atención en el estudio de los errores que giran en torno a la comprensión de ciertos conceptos de inferencia estadística y que se resumen en varios trabajos de síntesis (Batanero, 2020; Batanero et al., 2020; Castro-Sotos et al., 2007; de la Fuente y Díaz, 2004; Harradine, Batanero y Rossman, 2011). Asimismo, existe una gran controversia sobre la aplicación de la inferencia debido a los errores en la aplicación y la interpretación de la inferencia estadística en artículos de investigación del campo de la psicología o de educación, entre otros (Batanero, 2000; Caperos y Pardo, 2013; Frías et al., 2002, Díaz- Batanero et al., 2019).

Una posible causa de estos errores puede estar asociada al hecho de que la inferencia estadística es presentada, generalmente, bajo el marco frecuencial, como si hubiese una metodología única de la inferencia, sin tener en cuenta otras tendencias existentes en la práctica estadística (por ejemplo, la metodología bayesiana, fiducial o

remuestreo) que en la enseñanza o en la práctica se omiten o se confunden con la frecuencial (Gigerenzer, 1993; Harlow, Mulaik y Steiger, 2016) y que hemos analizado en un trabajo previo, para el caso del contraste de hipótesis (Batanero et al., 2017). Dicho marco frecuencial requiere la comprensión de numerosos conceptos y procedimientos que los estudiantes no siempre han asimilado correctamente.

En la actualidad, y salvo algunas nociones de muestreo en la Educación Secundaria Obligatoria y Bachillerato, la enseñanza de la inferencia estadística se contempla por primera vez en segundo curso de Bachillerato de la modalidad de Humanidades y Ciencias Sociales, extendiéndose posteriormente a la mayoría de las carreras universitarias y estudios de postgrado. igualmente, las pruebas de evaluación de bachillerato para acceso a la universidad y de admisión a la universidad (nombrada como PEvAU, EVAU o EBAU según la comunidad autónoma) integran, en la asignatura de Matemáticas Aplicadas a las Ciencias Sociales II, un problema de inferencia. Sin embargo, no se encuentran en los antecedentes análisis detallados del contenido de inferencia estadística de estas pruebas, que pongan de relieve la dificultad de los problemas propuestos. Para completar este hueco, una primera parte de la investigación que se desarrolla en esta memoria se orienta en realizar un estudio empírico y caracterización de la presencia de la inferencia estadística en el currículo de Bachillerato y de los problemas propuestos en las pruebas de acceso a la universidad, determinando las principales variables que los caracterizan.

Una adecuada enseñanza de la inferencia requiere de la preparación de los futuros profesores que deben enseñar el tema a sus estudiantes y cuyo conocimiento sobre inferencia apenas ha sido tenido en cuenta en la investigación previa. Por este motivo, una segunda parte de la memoria es un estudio de evaluación orientado a evaluar algunos elementos del conocimiento didáctico-matemático (Godino, 2009; Godino et al., 2017; Pino-Fan y Godino, 2015) para la enseñanza de la inferencia estadística de futuros profesores de Educación Secundaria Obligatoria y Bachillerato. Todo ello surge para iniciar una investigación cuya finalidad será proponer acciones formativas dirigidas a mejorar la formación de los profesores para la enseñanza de la inferencia estadística.

La memoria se organiza en torno a cinco capítulos, permitiendo los dos primeros contextualizar el tema de investigación, para posteriormente, en el tercero y cuarto,

presentar los dos estudios centrales de esta tesis. Toda la investigación realizada se cierra con el capítulo de conclusiones. A continuación, se sintetizan los contenidos de cada uno de estos capítulos.

En el Capítulo 1 se expone el planteamiento del problema, justificando el interés del tema elegido y analizando cómo está contemplado en el currículo de Bachillerato. Dada la importancia que tienen las pruebas de evaluación para el acceso a la universidad en esta etapa educativa, se realiza un estudio de los contenidos de inferencia estadística evaluables en dichas pruebas y se lleva a cabo un estudio histórico de las modificaciones realizadas en las directrices curriculares que la rigen. El análisis en profundidad del significado institucional de los objetos matemáticos relacionados con la inferencia estadística nos permitirá identificar cierta información sobre la comprensión que puede tener el estudiante. Asimismo, se presenta el marco teórico, enfoque ontosemiótico (EOS), empleado en esta memoria de tesis. Para finalizar se describen los objetivos sobre los que se ha desarrollado la investigación e hipótesis o expectativas iniciales que se espera encontrar, así como la organización del proceso de investigación seguido.

Con objeto de fundamentar teóricamente la investigación e intentar analizar el porqué de los fracasos que se producen en torno a la inferencia estadística, el Capítulo 2 recoge una revisión bibliográfica sobre los trabajos de investigación centrados en el estudio de la comprensión y dificultades que presentan los estudiantes cuando abordan dichos contenidos. Concretamente, estos trabajos se organizan en torno al contraste de hipótesis y otros conceptos de inferencia. Asimismo, se han integrado investigaciones focalizadas tanto en libros de texto como en las pruebas de acceso a la universidad y análisis de otras pruebas de evaluación.

El Capítulo 3 presenta el primero de los dos estudios empíricos que constituyen la tesis. Mediante un análisis de contenido se han clasificado los problemas propuestos en Andalucía en las pruebas de acceso a la universidad de Matemáticas Aplicadas a las Ciencias Sociales II, de acuerdo a diferentes campos de problema diferenciados en inferencia estadística. Para cada uno de los campos de problema, se realiza el análisis semiótico de la resolución de uno de ellos, identificando las prácticas y configuraciones de objetos matemáticos relacionados. Asimismo, se ha analizado el modelo de distribución de la población de partida en el problema; el parámetro que se pide estimar

y el contexto en el que se presenta el problema, dentro de los considerados en las pruebas PISA. Para cada una de las variables se ha realizado un estudio de la distribución de las diferentes categorías, globalmente y a lo largo del periodo seleccionado. A partir del estudio realizado, se lleva a cabo una reflexión sobre algunos componentes de la idoneidad didáctica de las pruebas de acceso a la universidad.

El Capítulo 4 resume un estudio de evaluación de algunos elementos del conocimiento didáctico-matemático de un grupo de futuros docentes de Educación Secundaria Obligatoria y Bachillerato sobre inferencia estadística. Para tal fin, se ha diseñado un cuestionario que ha permitido describir los conocimientos matemáticos de los futuros profesores en relación al contraste de hipótesis y al intervalo de confianza. A partir del análisis de las respuestas, se han identificado las dificultades y los conocimientos matemáticos que poseen los participantes respecto al tema en cuestión.

El estudio se completa analizando la componente cognitiva del conocimiento didáctico-matemático que poseen los futuros docentes realizando, para ello, una evaluación de los conocimientos que estos poseen en relación a los posibles errores que pueden cometer los estudiantes de Bachillerato cuando trabajan en el aula dichos contenidos. Igualmente, se analiza una pregunta enfocada en conocer cómo ellos, como futuros docentes, explicarían a los estudiantes de Bachillerato el intervalo de confianza, dándonos información sobre una parte de la componente mediacional de su conocimiento didáctico-matemático.

El Capítulo 5 recoge las conclusiones más destacadas de la memoria de tesis. Se recuperan los objetivos e hipótesis planteadas en el Capítulo 1 con objeto de analizar el cumplimiento de estos en base al proceso de investigación seguido. Asimismo, se destacan las principales aportaciones de la memoria dentro del campo de la didáctica de la estadística y las limitaciones del estudio. Para cerrar el capítulo se identifican las futuras líneas de trabajo que pueden enriquecer y complementar la investigación aquí presentada. La memoria se complementa con las referencias utilizadas y una lista de las publicaciones en que se ha desarrollado el contenido de la investigación.

Capítulo 1

PLANTEAMIENTO DEL PROBLEMA

Índice de Contenidos

- 1.1. Introducción
 - 1.2. Inferencia estadística y su importancia
 - 1.2.1. Desarrollo de la inferencia y su papel en el avance de las ciencias
 - 1.2.2. Importancia en la formación del estudiante
 - 1.3. La inferencia estadística en el currículo de Bachillerato
 - 1.4. Pruebas de evaluación del bachillerato para acceso a la universidad y de admisión a la Universidad
 - 1.4.1. Las pruebas de acceso a la universidad
 - 1.5. Marco teórico
 - 1.5.1. Tipos de objetos matemáticos
 - 1.5.2. Significados institucionales y personales y sus tipos
 - 1.5.3. Comprensión y evaluación en el EOS
 - 1.5.4. Función semiótica y conflicto semiótico
 - 1.5.5. Idoneidad didáctica y sus componentes
 - 1.5.6. Modelo del conocimiento del profesor en el EOS
 - 1.6. Objetivos e hipótesis de la investigación
 - 1.6.1. Objetivos del trabajo
 - 1.6.2. Hipótesis o expectativas iniciales
 - 1.7. Organización de la investigación
-

1.1. INTRODUCCIÓN

Como se ha indicado en la introducción a la tesis, la memoria de investigación que aquí se presenta tiene como objetivo principal analizar la inferencia estadística desde dos

puntos de vista: a) caracterizar los problemas de inferencia estadística propuestos en Andalucía en las pruebas de acceso a la universidad para las Matemáticas Aplicadas a las Ciencias Sociales II y b) evaluar algunos elementos de los conocimientos didáctico-matemáticos que poseen los futuros profesores de Matemáticas de Bachillerato sobre inferencia estadística.

El interés de este tema de investigación proviene fundamentalmente de la importancia que tiene la inferencia estadística y de mi propia experiencia como profesora en asignaturas del Área de Ciencias Sociales. Con objeto de contextualizar la investigación, se comienza el primer capítulo analizando el papel que ha tenido la inferencia estadística en el desarrollo de otras ciencias, así como la importancia que esta materia tiene en la formación de los estudiantes.

Para situar la enseñanza de la inferencia estadística en Bachillerato y en las pruebas de acceso a la universidad, se incluye un análisis de cómo las normativas curriculares abordan este tema y las modificaciones que se han producido en los últimos años. Seguidamente, se presenta el marco teórico que sustenta la investigación y que posteriormente permite describir los objetivos y formular algunas hipótesis, que se entiende cómo expectativas iniciales. Por último, y con objeto de facilitar la lectura de la presente memoria se incorpora una descripción de la organización del trabajo y de sus principales características metodológicas.

1.2. INFERENCIA ESTADÍSTICA Y SU IMPORTANCIA

Durante buena parte de la historia de las matemáticas, la estadística se ha reducido fundamentalmente a la estadística descriptiva, cuya principal tarea ha estado centrada en la recopilación de datos, clasificación de estos, tabulación, representación y análisis descriptivo. Esta concepción tan limitada de la estadística sufrió un cambio radical a partir de los años 30 del siglo XX con el nacimiento de la estadística inductiva o inferencia estadística, enfocada en la búsqueda de métodos que permitieran justificar las investigaciones realizadas en las ciencias empíricas (Hacking, 2006, Hald, 2008).

La inferencia estadística es una herramienta que posibilita la obtención de conclusiones de una población a partir de la información proporcionada por una

muestra representativa de la misma. Este hecho ha provocado que instituciones tales como *American Psychological Association* recomienden el uso de distintos métodos de la inferencia estadística (APA, 2010; Wilkinson, 1999).

En la sociedad actual la estadística es una herramienta ampliamente utilizada para describir, entre otros, datos económicos, políticos, sociales e incluso para analizar el curso de la pandemia debida a la COVID-19. Estamos inmersos en un mundo en el que existe una mayor necesidad de traducir todo en cifras para posteriormente analizarlas e interpretarlas. Ello lleva a la necesidad de conseguir una cultura estadística para todos los ciudadanos (Gal, 2002; Sharma, 2017), es decir, que se adquiriera la capacidad de leer e interpretar críticamente la información estadística y de utilizar esta información para apoyar argumentos y decisiones.

Los medios de comunicación, las redes sociales y prensa bombardean constantemente al ciudadano con argumentos y afirmaciones que son obtenidas, en su mayoría, mediante técnicas estadísticas inferenciales. Sin embargo, aunque el acceso a la información hoy en día no presenta gran complejidad, el razonamiento inferencial y como consecuencia, el análisis crítico, no está al alcance de cualquiera (Castro-Sotos et al., 2007), dejando entrever la necesidad de realizar acciones formativas al respecto.

Ante la necesidad de que los ciudadanos posean ciertos conocimientos y habilidades que posibiliten evaluar críticamente dicha información es fundamental formar a ciudadanos con base estadística que le permita entender correctamente la información recibida. Garfield y Ben-Zvi (2008) señalan que:

Hacer inferencias a partir de los datos es ahora parte de la vida cotidiana y la revisión crítica de los resultados de inferencia estadística a partir de estudios de investigación es una capacidad importante para todos los adultos, (p. 262).

Desde un punto de vista académico, aunque la enseñanza de la estadística y la probabilidad ha estado recogida en las normativas curriculares desde hace más de treinta años, las investigaciones revelan la existencia de errores sobre su comprensión y su aplicación (Batanero, 2019; Batanero et al., 2020; Castro-Sotos et al., 2007) y que su enseñanza se centra en mostrar conceptos, reglas y procedimientos alejados del mundo real (Bakker y Derry, 2011). No obstante, Batanero, Arteaga y Contreras (2011) destacan la tendencia de introducir la estadística y probabilidad desde edades

tempranas de una forma más experimental con objeto de garantizar un proceso de enseñanza-aprendizaje significativo. El marco de evaluación de PISA (OCDE, 2015) reconoce también la necesidad de que los estudiantes desarrollen las capacidades estadísticas necesarias para enfrentarse, en su vida diaria y profesional, a la toma de decisiones, tanto desde una perspectiva matemática como científica, en situaciones en las que se carece de información.

A lo largo de los siguientes puntos se analiza el papel que juega la inferencia estadística en el desarrollo de otras ciencias y en la formación del estudiante.

1.2.1. Desarrollo de la inferencia y su papel en el avance de las ciencias

La inferencia estadística es relativamente reciente, puesto que los primeros comienzos se remontan al siglo XVIII, cuando los matemáticos Thomas Bayes y Pierre Simon de Laplace dieron respuesta, de forma independiente, al *problema de la probabilidad inversa* correspondiendo a lo que se conoce como metodología bayesiana (Hald, 2008). Según la investigación realizada por Hald (2008), el autor señala que Laplace también fue pionero en el desarrollo del test de significación para la media para una distribución uniforme. Asimismo, Laplace (1749/1827) y Gauss (1777/1855) mostraron que la distribución aproximada de las medias muestrales para muestras grandes corresponde a una distribución normal.

Como ha sido señalado anteriormente, la inferencia estadística surge con el objetivos de buscar métodos que permitan justificar el razonamiento inductivo empírico y la extensión de sus conclusiones, problema de gran importancia en las ciencias empíricas (Batanero, Díaz-Batanero y López- Martín, 2017; Cabriá, 1994; Mayo y Cox, 2006; Rivadulla, 1991). A continuación, se resume esta problemática, partiendo del análisis previo recogido en Batanero et al. (2017). En el razonamiento inductivo empírico, la conclusión es más amplia que las premisas, y puede ser falsa, aunque se parta de premisas ciertas. La justificación de la validez de estas conclusiones ha sido motivo de discusión en filosofía y estadística, sin que se haya obtenido una solución aceptada por consenso (Black, 1979; Cabriá, 1994; de la Fuente y Díaz, 2004; Vallecillos, 1994). Dicho problema fue estudiado desde Aristóteles, aunque fue Hume el primero que los planteó con claridad en 1739, argumentando que ningún conjunto de

observaciones pasadas puede justificar sucesos que ocurran en el futuro (Hacking, 2006; Vallecillos, 1994).

Popper (1902/1994), en su libro *The logic of scientific discovery*, escrito en 1934 y vuelto a reescribir en 1959 propuso aceptar una teoría como provisionalmente cierta si, a pesar de numerosos intentos, no se consigue refutarla. El autor sugirió poner a prueba las hipótesis científicas, mediante experimentos u observaciones y comparar los patrones deducidos de la teoría con los datos obtenidos en dichos experimentos. La teoría sería provisionalmente confirmada si los datos recogidos siguiesen estos patrones y sería refutada si los datos la contradijeran. Es decir, no tiene la misma validez la refutación que la aceptación de una teoría (que siempre sería provisional).

Estas ideas tuvieron una gran repercusión en el desarrollo de la inferencia estadística. En particular, la inferencia frecuencial fue sistematizada entre otros autores por Fisher, Neyman y Pearson, quienes desarrollan diferentes teorías sobre los contrastes de hipótesis (Batanero y Borovcnik, 2016). Sus aproximaciones, aunque no se diferencian en lo que concierne a los cálculos, si tienen diferentes objetivos y se basan en diferentes conceptos.

- a. Fisher (1935), en su libro *The design of experiments*, propone el *test de significación* como un procedimiento el cual que posibilita el rechazo de una hipótesis determinada, en función de los datos obtenidos en un experimento y con un nivel de significación determinado.
- b. Neyman y Pearson (1933) por su parte estaban interesados en el contraste de hipótesis como un proceso de decisión que permite elegir entre una hipótesis dada H_0 y otra hipótesis alternativa H_1 (Rivadulla, 1991). Además, los autores introducen la idea de error tipo II (probabilidad de no rechazar la hipótesis nula sabiendo que es falsa), no tenida en cuenta por Fisher.

Además del enfoque frecuencial se desarrollaron otras metodologías. En particular destacamos la teoría bayesiana, porque algunos estudiantes interpretan conceptos propios de la inferencia frecuencial en forma bayesiana. En esta metodología, que es, en realidad, anterior a la frecuencial (Batanero y Borovcnik, 2016) tiene como finalidad actualizar la probabilidad de una hipótesis, entendida como el grado de creencia personal, que va desde la falsedad absoluta (cero) y la máxima certeza absoluta

(uno) (Gigerenzer, 1993; Lecoutre, Lecoutre y Poitevineau, 2001).

Aunque la metodología frecuencial y la bayesiana tienen en común la recolección de datos como base de la estimación de parámetros desconocidos, en la práctica la principal diferencia se basa en el tratamiento de dichos parámetros. Mientras que en el enfoque frecuencial el parámetro es considerado como un valor fijo y desconocido, el enfoque bayesiano interpreta el parámetro como una variable aleatoria cuya distribución de probabilidad inicial es actualizada en base al teorema de Bayes para obtener una distribución final, que se utiliza para contrastar hipótesis o calcular intervalos de credibilidad para el parámetro (Díaz-Batanero, 2005; 2018).

Ambas metodologías, tal y como se ha comentado, tienen diferentes fines y son apropiadas en distintas situaciones. Dentro del ámbito académico, es importante que el futuro docente distinga y haga ver a sus estudiantes las dos aproximaciones para promover un uso adecuado de la inferencia estadística.

En la actualidad es indiscutible la influencia que tiene la estadística como disciplina en diversos campos de conocimiento dado que permite analizar situaciones en la que existe una fuerte componente aleatoria. Es indiscutible que el desarrollo de algunos campos de conocimiento no hubiera sido posible sin el uso de las distintas herramientas estadísticas, como, por ejemplo, la psicología (eficacia en psicoterapia, como afecta los niveles de estrés en el trabajo personal, etc.), la ingeniería (plantas de control de contaminación, diseño estructural de industriales, etc.), la economía (estudio de opinión, toma de decisiones, análisis de riesgo, procesos de manufacturación, etc.), la medicina (desarrollo de nuevos medicamentos o vacunas, etc.) y la política, entre otros. Es también incuestionable que el avance de la tecnología computacional ha jugado un papel esencial para fortalecerla y consolidarla en las últimas décadas.

1.2.2. Importancia en la formación del estudiante

El impacto adquirido por la inferencia estadística se ha visto reflejado también en las normativas curriculares incluyéndola en algunas de las etapas educativas para su enseñanza y aprendizaje. En la instrucción de los estudiantes de Bachillerato se considera que una buena formación en inferencia estadística capacita a cualquier

ciudadano para:

- Analizar, interpretar y valorar fenómenos sociales, con objeto de comprender los retos que plantea la sociedad actual;
- Formar criterios propios sobre fenómenos sociales y económicos;
- Interpretar datos y mensajes, argumentando con precisión y rigor;
- Formular hipótesis que permiten contrastar estrategias para la resolución de problemas;
- Justificar procedimientos aportando rigor a los razonamientos, detectando inconsistencias lógicas;
- Adquirir y manejar con fluidez un vocabulario específico de términos y notaciones matemáticos;
- Establecer relaciones entre las matemáticas y el entorno social, cultural o económico.

La mayoría de los estudiantes de la etapa de Bachillerato seguirán estudios de ciencias sociales, como economía, psicología, administración y dirección de empresas, periodismo o comunicación, entre otras. En todas ellas es necesaria una buena base en inferencia, tanto para la lectura de artículos en revistas de su especialidad, como para llevar a cabo sus futuras tareas profesionales, hecho contemplado en la elaboración de los títulos de grado en los que se incluyen materias sobre inferencia estadística. En consecuencia, será necesario que los estudiantes en su etapa de Bachillerato adquieran una buena base en la parte elemental de la inferencia estadística para afrontarla en el ámbito universitario.

Una adecuada formación en inferencia estadística contribuye igualmente a formar a ciudadanos autónomos que pueden dar respuesta a muchas de las necesidades que requiere la sociedad en la que vivimos actualmente. Estos conocimientos formarían parte de la cultura estadística (Gal, 2002; Sharma, 2017) descrita anteriormente, ya que con más frecuencia los medios de comunicación incluyen estudios de muestreo, encuestas o resultados de investigaciones basados en métodos de inferencia. Esta necesidad se ha visto reflejada en el currículo de Bachillerato, que

se comenta a continuación.

1.3. LA INFERENCIA ESTADÍSTICA EN EL CURRÍCULO DE BACHILLERATO

En España en la actualidad la estructura de Bachillerato y sus enseñanzas mínimas se rigen según la Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa (LOMCE) que modifica la Ley Orgánica de Educación 2/2006, de 3 de mayo (LOE). La LOMCE es la séptima ley orgánica de la democracia española y fue implantada únicamente con los votos a favor del Partido Popular. Desde su publicación, la implantación de la ley ha encontrado muchos obstáculos, ya que algunas comunidades autónomas mostraban su rechazo a la implantación de la reforma educativa.

En la comunidad andaluza y en base al Decreto 97/2015, de 3 de marzo, se establece la ordenación y las enseñanzas correspondientes a la educación primaria en Andalucía, implantando todos los cursos de educación primaria (CECD, 2015). En el caso de la Educación Secundaria Obligatoria y del Bachillerato la situación es totalmente diferente. Puesto que el Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de ambas etapas, se hizo público el día 3 de enero de 2015, no fue posible elaborar y aprobar la normativa con los nuevos aspectos de la LOMCE en la comunidad de Andalucía. Por ello, se decidió que el curso académico 2015/2016 fuera un curso transitorio, mientras se aprobaba la normativa autonómica de adaptación a la LOMCE. Durante el mencionado curso académico se implantó únicamente los cursos impares de Educación Secundaria Obligatoria (1º y 3º ESO) y de Bachillerato (1º Bachillerato) y al año siguiente se implantaron los cursos restantes.

En los estudios de Bachillerato se distinguen tres modalidades, Ciencias y Tecnología, Humanidades y Ciencias Sociales y Artes, siendo en las dos primeras donde se incluye la materia de Matemáticas. Concretamente en la primera modalidad las asignaturas de primero y segundo son Matemáticas I y Matemáticas II, respectivamente. En la modalidad de Humanidades y Ciencias Sociales las materias reciben el nombre de Matemáticas Aplicadas a las Ciencias Sociales I, en primero, y Matemáticas Aplicadas a las Ciencias Sociales II, en segundo.

El cambio de la LOE a la LOMCE supuso una transformación en la asignatura Matemáticas II ya que pasó de tres bloques de contenidos (Álgebra Lineal, Geometría y Análisis) a cuatro, con la incorporación del Bloque de Estadística y Probabilidad. Sin embargo, aunque en su estructura las asignaturas de Matemáticas de la modalidad de Tecnología integran el Bloque de Estadística y Probabilidad, se diferencian en los contenidos a tratar, en general, los contenidos de Matemáticas II son los que con la LOE se trabajaban en Matemáticas I.

Centrando la atención en los contenidos de inferencia estadística correspondientes a la asignatura de Matemáticas de la modalidad de Humanidades y Ciencias Sociales, en la Tabla 1.3.1 recoge los contenidos correspondientes al mencionado bloque. Obsérvese que, aunque coincide en un alto porcentaje los contenidos de inferencia estadística en ambas normativas curriculares, se destaca que en el caso de la LOMCE son especificados con mayor detalle. Por ejemplo, en el caso de los intervalos de confianza, además de incluir la construcción del propio intervalo, la LOMCE aborda el estudio de la relación existente entre confianza, error y tamaño muestral. Sin embargo, un hecho destacable es que el inicio de la LOMCE supuso la exclusión de los contrastes de hipótesis.

Los contenidos recogidos en la Tabla 1.3.1. abarcan prácticamente un curso elemental de inferencia estadística que se pueden encontrar en los primeros cursos de grado. Como novedad, la nueva regulación (MECD, 2015b) incluye, junto a los criterios de evaluación (criterios para evaluar, tanto en conocimientos como en competencias, el aprendizaje del alumnado), los estándares de aprendizajes evaluables en todas las asignaturas, definidos como

Especificaciones de los criterios de evaluación que permiten definir los resultados de aprendizaje, y que concretan lo que el estudiante debe saber, comprender y saber hacer en cada asignatura; deben ser observables, medibles y evaluables y permitir graduar el rendimiento o logro alcanzado. Su diseño debe contribuir y facilitar el diseño de pruebas estandarizadas y comparables. (p. 172).

Puesto que los criterios de evaluación como los estándares de aprendizaje evaluables son esenciales en la planificación del currículo, en el Tabla 1.3.2 se muestran los relacionados con el tema de inferencia estadística.

El análisis de los contenidos de Inferencia estadística de las normativas curriculares nos permite ver el cambio de tendencia de formar a ciudadanos estadísticamente cultos (Batanero, Arteaga y Gea, 2012). Al finalizar el segundo curso de Bachillerato el estudiantado debe poseer suficientes recursos que le permitan analizar críticamente la información recibida desde los distintos medios de comunicación o desde cualquier ámbito.

Tabla 1.3.1.

Contenidos del Bloque 4 de Estadística y Probabilidad tratados en la LOE (MEC, 2007, p.45476) y la LOMCE (MECD, 2015b, p.389)

LOE	LOMCE
Implicaciones prácticas de los teoremas: Central del límite, de aproximación de la Binomial a la Normal y Ley de los Grandes Números.	Población y muestra. Métodos de selección de una muestra. Tamaño y representatividad de una muestra.
Problemas relacionados con la elección de las muestras. Condiciones de representatividad. Parámetros de una población.	Estadística paramétrica. Parámetros de una población y estadísticos obtenidos a partir de una muestra. Estimación puntual.
Distribuciones de probabilidad de las medias y proporciones muestrales.	Media y desviación típica de la media muestral y de la proporción muestral. Distribución de la media muestral en una población normal. Distribución de la media muestral y de la proporción muestral en el caso de muestras grandes.
Intervalo de confianza para el parámetro p de una distribución binomial y para la media de una distribución normal de desviación típica conocida.	Estimación por intervalos de confianza. Relación entre confianza, error y tamaño muestral.
Contraste de hipótesis para la proporción de una distribución binomial y para la media o diferencias de medias de distribuciones normales con desviación típica conocida.	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.

A pesar de la transformación de los documentos curriculares, desde un punto de vista didáctico, la inclusión de este contenido supone un desafío para el profesorado debido a la existencia de una gran variedad de dificultades y errores por parte del alumnado y por parte de algunos profesores encargados de su enseñanza (Castro-Sotos et al., 2007; Harradine et al., 2011, entre otros).

Tabla 1.3.2.*Criterios de Evaluación y Estándares de Aprendizaje Evaluables MECD (2015b, p.389)*

Criterios de evaluación	Estándares de aprendizaje evaluables
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	<p>2.1. Valora la representatividad de una muestra a partir de su proceso de selección</p> <p>2.2. Calcula estimadores puntuales para la media, varianza, desviación típica y proporción poblacionales, y lo aplica a problemas reales</p> <p>2.3. Calcula probabilidades asociadas a la distribución de la media muestral y de la proporción muestral, aproximándolas por la distribución normal de parámetros adecuados a cada situación, y lo aplica a problemas de situaciones reales</p> <p>2.4. Construye, en contextos reales, un intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida</p> <p>2.5. Construye, en contextos reales, un intervalo de confianza para la media poblacional y para la proporción en el caso de muestras grandes</p> <p>2.6. 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</p> <p>3.1. Utiliza las herramientas necesarias para estimar parámetros desconocidos de una población y presentar las inferencias obtenidas mediante un vocabulario y representaciones adecuadas</p> <p>3.2. Identifica y analiza los elementos de una ficha técnica en un estudio estadístico sencillo</p> <p>3.3. Analiza de forma crítica y argumentada información estadística presente en los medios de comunicación y otros ámbitos de la vida cotidiana</p>

1.4. PRUEBAS DE EVALUACIÓN DEL BACHILLERATO PARA ACCESO A LA UNIVERSIDAD

Las pruebas de acceso a la universidad consisten en una prueba de madurez que permite evaluar los conocimientos y capacidades adquiridas por los futuros universitarios durante su etapa de Bachillerato (primer y segundo curso). Dichas pruebas evalúan la competencia lingüística, la competencia matemática y las competencias básicas en ciencia y tecnología, entre otras. Para estudiar la implantación de estas hay que remontarse a la década de los cuarenta (Muñoz Vitoria, 1995).

En el trabajo presentado por Muñoz Vitoria (1995) sobre una perspectiva histórica de las pruebas de acceso a la universidad, el autor distingue tres etapas según la denominación y el enfoque que adquiere la prueba.

- Primera etapa, desde 1940 hasta 1953. Por la Ley de Reforma de la segunda enseñanza del 20 de septiembre de 1938, se instaura el examen de estado. La realización del examen se lleva a cabo al finalizar el séptimo curso.
- Segunda etapa, desde 1953 hasta 1971. La prueba recibe el nombre de Prueba de Madurez y se implanta por la Ley sobre ordenación de la Enseñanza Media del 26 de febrero de 1953. La prueba de madurez la realizaban los estudiantes que habían cursado el conocido como curso preuniversitario.
- Tercera etapa, desde 1971 hasta 1990¹. En 1970 se publica la Ley General de Educación y Financiamiento de la Reforma Educativa. Sin embargo, a lo largo de este período el cambio más significativo, con respecto a la prueba, se tiene en 1975. Con la Ley 30/1974 se instaura lo que se conoce como las pruebas de aptitud para el acceso a la universidad (PAAU) cuyo diseño tenía como finalidad reconocer a todos aquellos alumnos que estaban capacitados para iniciar dichos estudios (universitarios). En esta época la prueba estaba destinada a los estudiantes que realizan el curso de orientación universitaria (COU).

Las pruebas de aptitud para el acceso a la universidad han sido uno de los formatos más estables, manteniéndose hasta 1990. Durante ese año, se promulga la Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo (LOGSE) derogando la Ley General de Educación de 1970. La implantación de la LOGSE provocó un replanteamiento de las pruebas de acceso a la universidad ya que se suprime el curso de orientación y preparación para el acceso a la universidad (conocido como COU) y las directrices específicas para preparar a los alumnos, del segundo ciclo de enseñanza Secundaria, que deseaban ingresar en la universidad. Este supuesto estado de estabilidad se difuminó pues desde entonces se han realizado seis modificaciones.

¹ En los periodos 1971-1972 y 1973-1974 se eliminaron las pruebas de acceso.

La Tabla 1.4.1 recoge las modificaciones realizadas desde el punto de vista legislativo desde 1999. En la página oficial del Ministerio de Educación, Ciencia y Deporte es posible encontrar seis reales decretos (dos de ellas son modificaciones sobre el Real Decreto 1640/1999 y 990/2000) cuyo objetivo son la regulación de las pruebas de acceso a la universidad.

Tabla 1.4.1.

Reales Decretos que Regulan las Pruebas De Acceso a la Universidad Aprobados en el Periodo 1999-2014

Año	Real Decreto
1999	Real Decreto 1640/1999, de 22 de octubre, por el que se regula la prueba de acceso a estudios universitarios
2000	Real Decreto 990/2000, de 2 de junio, por el que se modifica y completa el Real Decreto 1640/1999, de 22 de octubre, por el que se regula la prueba de acceso a estudios universitarios
2002	Real Decreto 1025/2002, de 4 de octubre, por el que se modifica el Real Decreto 1640/1999, de 22 de octubre, modificado y completado por el Real Decreto 990/2000, de 2 de junio, por el que se regula la prueba de acceso a estudios universitarios
2003	Real Decreto 1741/2003, de 19 de diciembre, por el que se regula la prueba general de Bachillerato
2008	Real Decreto 1892/2008, de 14 de noviembre, por el que se regula las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades públicas españolas
2014	Real Decreto 412/2014, de 6 de junio, por el que se establece la normativa básica de los procedimientos de admisión a las enseñanzas universitarias oficiales de Grado

A lo largo de la presente sección se centrará el estudio en realizar una breve revisión de las pruebas de acceso a la universidad regidas en los Reales Decretos que se han aprobado en las dos últimas décadas.

1.4.1. Las pruebas de acceso a la universidad

El establecimiento de la LOMCE supuso la modificación de los requisitos de acceso y admisión a las enseñanzas oficiales de Grado eliminando la superación de la prueba de acceso como requisito de acceso a las enseñanzas universitarias, siendo las propias universidades las encargadas de determinar la admisión a estas enseñanzas (MECD, 2014). En este escenario, entre las distintas formas de admisión a las enseñanzas universitarias se establece que las universidades pueden determinar la admisión considerando exclusivamente la calificación final de Bachillerato o fijando un

procedimiento determinado. Bajo el supuesto de fijar un procedimiento de acceso se utiliza alguno o algunos de los siguientes criterios (MECD, 2014, p. 43313):

- a) Modalidad y materias cursadas en los estudios previos equivalentes al Título de Bachiller, en relación con la titulación elegida.
- b) Calificaciones obtenidas en materias concretas cursadas en los cursos equivalentes al Bachillerato español, o de la evaluación final de los cursos equivalentes al de Bachillerato español.
- c) Formación académica o profesional complementaria.
- d) Estudios superiores cursados con anterioridad.

El Proyecto de Real Decreto por el que se regulan las características generales de las pruebas de la evaluación final de Educación Primaria, y las características de las pruebas de las evaluaciones finales de Educación Secundaria Obligatoria y de Bachillerato presentado el 24 de marzo de 2015 destaca las características generales de la prueba de la evaluación final de Bachillerato:

La finalidad de la evaluación final de Bachillerato es comprobar el logro de los objetivos de la etapa y el grado de adquisición de las competencias correspondientes por los alumnos y alumnas en relación con las siguientes materias: a) Todas las materias generales cursadas en el bloque de asignaturas troncales. [...]. b) Dos materias de opción cursadas en el bloque de asignaturas troncales, en cualquiera de los cursos. [...]. c) Una materia del bloque de asignaturas específicas cursada en cualquiera de los cursos, que no sea Educación Física ni Religión. (MECD, 2015a, p.9).

Un estudio comparativo de la normativa actual y la establecida en la LOE nos permite ver que ambas son prácticamente iguales en relación a su finalidad, requisitos de participación, convocatorias, entre otros elementos. Desde la implantación de la LOE, las pruebas de acceso a la universidad experimentaron principalmente cambios en su estructura y puntuación, siendo posible alcanzar una calificación de 14 puntos. La incorporación de estas modificaciones estableció una relación más estrecha entre la citada prueba y los estudios universitarios que se pretenden cursar.

Entre los cambios introducidos, se destacó la desaparición de las vías de examen, siendo los propios estudiantes los que rediseñan la prueba, es decir, los alumnos eligen, las hayan cursado o no, las materias de la modalidad que presentan mayor afinidad con los estudios universitarios que desean cursar. El modelo descrito está determinado por

dos fases, una de ellas obligatoria para todos los estudiantes que realizan la prueba, denominada Fase General, y otra, de carácter optativo, Fase Específica, con la que se puede incrementar la calificación final de la prueba de acceso hasta 4 puntos. A continuación, se detalla cada fase:

- *Fase General.* El objetivo de esta fase consiste en la valoración de la madurez y destrezas básicas del estudiante. Consta de cuatro exámenes que se corresponden a las materias comunes de segundo de Bachillerato. Los estudiantes se examinan de una primera prueba relacionada con los contenidos de Lengua Castellana y Literatura II; una segunda prueba de Lengua Extranjera (elegir una entre alemán, francés, inglés, italiano o portugués) y una tercera prueba de Historia de España. Por último, la cuarta prueba corresponde a una materia de la modalidad del Bachillerato cursado. En el caso particular de las comunidades de Andalucía, Asturias y Canarias, el estudiante podrá elegir, en la cuarta prueba, cualquier materia independientemente de la modalidad de Bachillerato cursada.

La calificación de la Fase General es la media aritmética de las calificaciones numéricas obtenidas de cada uno de los exámenes, obteniéndose una calificación máxima de 10 puntos. La superación de esta fase permite que el estudiante pueda acceder a los estudios universitarios y la calificación obtenida en esta fase tiene una validez indefinida.

- *Fase Específica.* La finalidad de esta fase es la evaluación de los conocimientos en unos ámbitos disciplinares concretos relacionados con los estudios vinculados a la rama de conocimiento que se quiere cursar en la universidad. Por lo que, la realización de los ejercicios de esta fase permite obtener la nota de admisión para cada uno de los títulos de grado a los que el estudiante desea acceder. En esta fase el estudiante podrá examinarse de hasta un máximo de cuatro materias, tomándose las calificaciones de dos materias, con una calificación igual o superior a 5 puntos, que otorguen la mejor nota de admisión una vez ponderadas con un 10% o 20%. Si las materias están vinculadas a la rama del conocimiento de la titulación a la que desea acceder, el parámetro de ponderación de la materia es del 20% y en caso contrario será

del 10%. A diferencia de la Fase General, las calificaciones de las materias cursadas en esta fase tienen una validez de dos cursos académicos siguientes a la superación de la misma.

La nota de acceso a la universidad se obtiene como la media ponderada de la calificación de la Fase General multiplicada por un factor de ponderación del 40% y la calificación media del expediente multiplicada por un factor de ponderación del 60%, es decir,

$$\text{Nota Acceso} = 40\% \text{ Calificación F.G.} + 60\% \text{ Calificación Media Expediente.}$$

La superación de la prueba de acceso está condicionada a obtener en la Fase General una calificación igual o superior a 4 y a su vez, la calificación final de la nota de acceso deberá ser mayor o igual a 5 puntos.

Por otro lado, los estudiantes que hayan realizado la Fase Específica podrán optar a una calificación máxima de 14 puntos. En este caso, la calificación, nombrada como Nota de Admisión, se obtiene como:

$$\text{Nota de Admisión} = \text{Nota de Acceso} + a \times M_1 + b \times M_2,$$

donde M_1 y M_2 , son, de entre todas las materias examinadas en la Fase Específica, las mejores notas de admisión, una vez multiplicadas por los parámetros a y b , para cada una de las carreras solicitadas por el alumno. No obstante, teniendo en cuenta los criterios marcados por las distintas universidades, los factores de ponderación pueden incrementarse un 10%, es decir, si las materias cursadas están estrechamente relacionadas con los futuros estudios universitarios, entonces las universidades tendrán la potestad de considerar una ponderación del 20%. El incremento de los factores de ponderación está condicionado a la existencia de una estrecha afinidad entre la optativa seleccionada y la rama de conocimiento a la que está adscrito el título de grado que se desea realizar.

De forma habitual, en la comunidad andaluza cada año se celebran dos convocatorias de las pruebas de acceso a la universidad, en el mes de junio (convocatoria ordinaria) y en el mes de septiembre (convocatoria extraordinaria). La situación de crisis sanitaria de la COVID-19 ha implicado un cambio en el diseño de las pruebas y en el calendario de celebración de las pruebas de acceso. Como novedad la

convocatoria de junio se ha trasladado al mes de julio y gran parte de los exámenes se han diseñado con una única opción de selección, incluyendo más preguntas de selección permitiendo así descartar el temario que puede no haber sido impartido a lo largo del curso académico. En relación a la estructura y duración de las pruebas no se ha producido variaciones respecto a convocatorias anteriores. Esta forma de proceder por parte de las autoridades responsables ha estado motivada por el hecho de garantizar la equidad de todo el estudiantado independientemente de las condiciones en las que han cursado el tercer y último trimestre del curso 2019/20.

1.5. MARCO TEÓRICO

Para llevar a cabo la investigación planteada es importante partir de un marco teórico fuerte, que considere tanto el significado como la comprensión de un objeto matemático más allá de la simple definición del mismo y que considere también un modelo de los diferentes componentes del conocimiento del profesor de matemáticas.

Nos apoyamos en herramientas teóricas propuestas en el enfoque ontosemiótico (EOS) del conocimiento y la instrucción matemáticos, sistema teórico elaborado por Godino y sus colaboradores (Godino, 1996; 2002, 2017; Godino y Batanero, 1994; 1998; Godino, Batanero y Font; 2007; 2020) con objeto de poner a disposición de los investigadores del campo de la didáctica de la matemáticas diversos instrumentos que permitan analizar los componentes de los procesos de enseñanza y aprendizaje. Dicho marco teórico asume que las matemáticas provienen de la actividad humana orientada a la resolución de determinados tipos de problemas, los cuales constituyen la razón de ser y a partir de los cuáles se dota de significado a los objetos emergentes de la misma.

1.5.1. Tipos de objetos matemáticos

En la práctica matemática intervienen una gran variedad de objetos matemáticos (conceptos, proposiciones, teorías) que tienen carácter inmaterial, por lo que son representadas de forma textual, oral, gráfica, etc. En el EOS se parte de la idea de problema y práctica considerada ésta como toda actuación o expresión realizada por alguien para resolver problemas matemáticos o extra-matemáticos, comunicar a otros

la solución obtenida, validarla o generalizarla a otros contextos y problemas (Font, Godino y Gallardo, 2013; Godino y Batanero, 1994; 1998).

En dicho marco, se asume que los objetos matemáticos, en nuestro caso los objetos relacionados con la inferencia estadística surgen de las prácticas matemáticas realizadas para resolver problemas relacionados con el objeto dado. El sentido dado a este término por los autores es muy amplio y puede englobar cualquier tarea o ejercicio que requiera el uso de las matemáticas en su resolución (Godino, 2002; 2003). Los problemas de inferencia surgen en el aula (problemas propuestos a los estudiantes), pero también en diferentes ciencias, que deben realizar estimaciones de parámetros o contrastar hipótesis para justificar los resultados de sus investigaciones.

Ejemplos de problemas relacionados con la inferencia y propuestos en las pruebas de acceso a la universidad serían determinar el tamaño de muestra necesario para llevar a cabo una estimación por intervalos con una precisión dada o contrastar una hipótesis sobre la media de una población (Batanero y Borovcnick, 2016).

El EOS considera, en un primer nivel, una tipología de objetos (Godino, Batanero y Font, 2007; Godino, Font, Contreras y Wilhelmi, 2006) que usaremos en el análisis de los problemas de inferencia propuestos en las pruebas de acceso a la universidad y en la evaluación del conocimiento de los futuros profesores.

- Son *situaciones-problemas* cualquier aplicación matemática o extraña a la matemática de la que surge un objeto matemático. Un ejemplo en nuestro estudio sería estimar la media poblacional tomando para ello una muestra de datos de dicha población.
- Son elementos del *lenguaje* las palabras, símbolos, expresiones, representaciones gráficas, que son empleadas para visualizar los datos del problema o de las operaciones que hacemos con ellos. En el caso particular de la inferencia estadística serían ejemplos las palabras sesgo, intervalo de confianza, hipótesis, así como los símbolos asociados y sus representaciones gráficas.
- Son *conceptos* las generalizaciones que son utilizados implícita o explícitamente por el alumno cuando resuelve el problema. Entre otros, en el tema que nos ocupa encontramos el de variable aleatoria y estadística,

población y muestra, parámetro y estadístico, intervalo, amplitud, estimación o hipótesis.

- Son *proposiciones* las relaciones entre los objetos matemáticos. En el presente estudio un ejemplo de ello sería que el área bajo la curva normal es igual a la unidad.
- Son *procedimientos* son el conjunto de operaciones o algoritmos que permiten resolver un problema. En nuestro caso podemos destacar la obtención del percentil que corresponde a un valor crítico que determina la región de rechazo.
- Son *argumentos* las explicaciones utilizadas para justificar una propiedad o explicar la solución de un problema. Un ejemplo de ello sería la demostración informal del teorema central del límite.

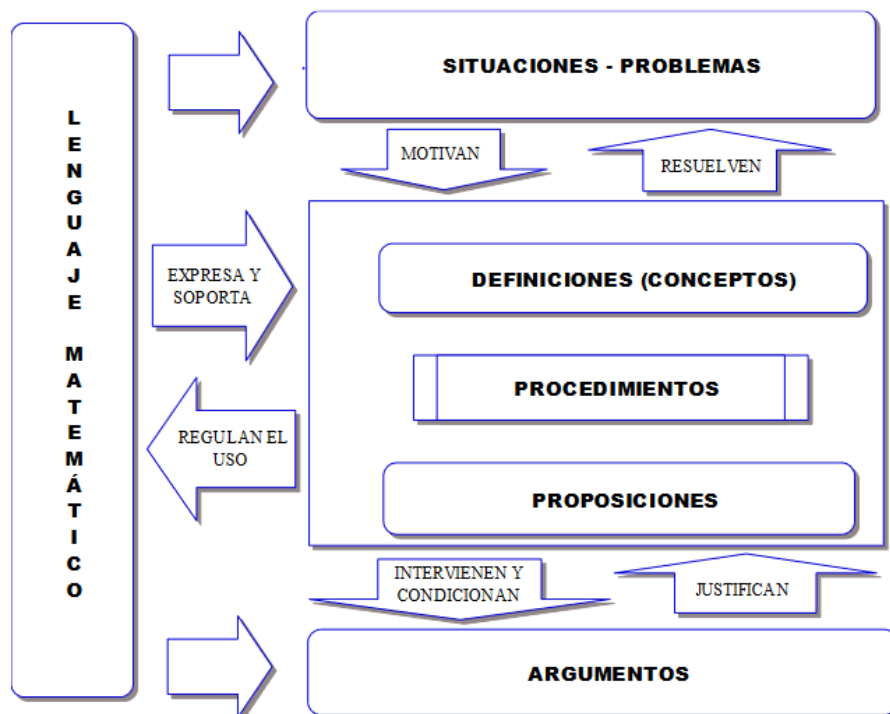
Las relaciones o redes que se establecen entre las distintas tipologías de objetos en la práctica matemática o en la enseñanza lleva a redes de objetos que el EOS denomina configuraciones, que pueden ser *epistémicas*, si son propias de la institución o *cognitivas*, si son específicas de una persona. En el análisis de los problemas propuestos en las pruebas de acceso a la universidad se describirán las configuraciones epistémicas ligadas a los diferentes campos de problemas identificados en el análisis. En la Figura 1.5.1 se muestra las relaciones entre los diferentes tipos de objetos de una configuración, que surgen en una práctica matemática.

1.5.2. Significados institucionales y personales y sus tipos

Godino (2002) examina diversas teorías sobre el significado y propone su propia visión sobre el mismo para estudiar la relación entre pensamiento, lenguaje y problemas. Indica que los símbolos representan objetos matemáticos y que el fin principal de la enseñanza no es solo conseguir que los estudiantes dominen la sintaxis del lenguaje simbólico, sino todo su significado.

Mediante la reiteración de las prácticas matemáticas realizadas al resolver muchos problemas similares, se llega a configurar el *significado del objeto matemático* o conjunto de prácticas que realiza una persona o se comparten en una institución para resolver un tipo de situaciones problema.

Figura 1.5.1. Configuración de Objetos Primarios



Nota: Tomado de Font y Godino (2006, p. 69).

Por tanto, los problemas propuestos a los estudiantes, bien durante la enseñanza o en las pruebas de evaluación van a incidir directamente en las prácticas matemáticas que realizan y, por consecuencia, en la adquisición de diferentes objetos matemáticos y el significado asignado a dichos objetos. De ello se deduce la importancia del análisis de tales problemas.

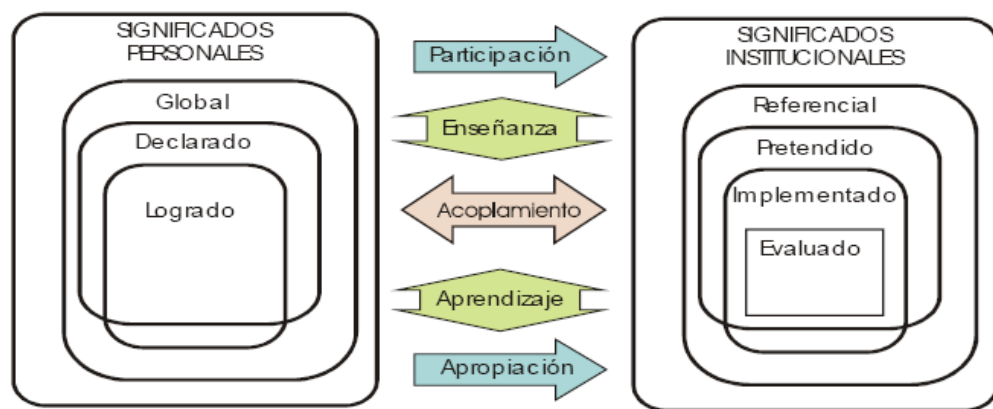
En el marco teórico EOS los autores diferencian entre el *significado institucional*, considerado como aquel que es compartido dentro de una institución, y el *significado personal*, siendo el adquirido por una persona, por ejemplo, un estudiante y puede ser diferente al aceptado dentro de la institución colaboradores (Godino, 2002; Godino et al., 2007; 2019) (véase Figura 1.5.2).

Dentro de la institución de enseñanza, se pueden distinguir cuatro tipos de significados:

- El significado *referencial* sería el más amplio pues es el empleado como referencia para elaborar una investigación o una acción en el aula.

Correspondería al conjunto más amplio de prácticas matemáticas relacionadas con el objeto de interés y se deduce del análisis de libros matemáticos sobre el tema, de la historia de su desarrollo, etc. En nuestro caso, este tipo de significado correspondería al significado de la inferencia frecuencial en la propia estadística.

Figura 1.5.2. *Tipos de Significados Institucionales y Personales*



Nota: Tomado de Godino (2002, p.141).

- El significado *pretendido* del objeto dentro de un nivel educativo sería el fijado por la institución, pues corresponde al sistema de prácticas incluidas en la planificación del proceso de enseñanza. Atendiendo a diversos factores de las orientaciones curriculares como, por ejemplo, el nivel educativo, la temporalización, conocimientos previos de los estudiantes, etc., se concreta el significado pretendido. En nuestro caso se considera como significado pretendido de la inferencia el recogido en las orientaciones curriculares para Bachillerato.
- El significado *implementado* es el que el profesor presenta a los estudiantes en cada centro o aula. Generalmente tiene algunas diferencias con el significado pretendido, porque la misma dinámica del aula hace que durante el proceso de enseñanza se produzcan pequeños cambios en el significado pretendido. En este trabajo no se considera este tipo de significado, pues no

se realiza la observación o el análisis de una secuencia de enseñanza.

- El significado *evaluado* corresponde al subsistema de prácticas que el propio docente utiliza con objeto de evaluar los aprendizajes de los estudiantes. En nuestro caso, ese subsistema de prácticas vendría determinado por las prácticas matemáticas requeridas en la resolución de los problemas planteados en las pruebas de acceso a la universidad.

De igual modo dentro de los significados personales de un objeto matemático se identifican tres tipos:

- El *significado global* que corresponde con todo el sistema potencial de prácticas subjetivas que es capaz de manifestar el estudiante en relación al objeto matemático en cuestión. Como es muy amplio, usualmente se restringe a una parte del mismo, a efecto de investigación.
- El *significado declarado* está asociado a las prácticas recogidas en las pruebas de evaluación, en nuestro caso, corresponden a la respuesta dada por el estudiante en la prueba de acceso a la universidad. Este significado considera las respuestas dadas correctamente y aquellas que desde la institución son incorrectas en una tarea o en una prueba de evaluación. En este trabajo, para el Estudio 2, sería significado declarado el conjunto de prácticas matemáticas que se puedan deducir de las respuestas de los futuros profesores a las tareas propuestas.
- El *significado logrado* relacionado con las prácticas que coinciden con las pretendidas institucionalmente. Sería aquella parte del significado declarado que se considere correcto en el estudio.

1.5.3. Comprensión y evaluación en el EOS

Como ha sido mencionado anteriormente, uno de los estudios recogidos en la presente memoria se orienta en determinar el significado institucional de la inferencia estadística evaluado en las pruebas de acceso de la materia Matemáticas Aplicadas a las Ciencias Sociales II. La existencia de un apropiado vínculo entre el significado institucional pretendido y el evaluado para un contenido, es primordial para asegurar, en nuestro caso, que las pruebas de evaluación sean coherentes con los contenidos fijados institucionalmente. Por tal motivo, en el Capítulo 3 se evalúa la adecuación de esa

correspondencia para el contenido pretendido de la inferencia en las orientaciones curriculares de Bachillerato.

Asimismo, se evalúan algunos elementos de los conocimientos didáctico-matemáticos sobre la inferencia por parte de los futuros profesores. La evaluación del conocimiento matemático de los futuros profesores llevada a cabo en el Estudio 2 permitirá deducir también su comprensión de la inferencia estadística.

Puesto que la comprensión de un determinado objeto matemático es un constructo psicológico inobservable, Godino (1996) estableció que el análisis de las respuestas de los estudiantes a los ítems, tareas o pruebas de evaluación puede permitirnos adquirir cierta información sobre la comprensión que tiene el estudiante, en nuestro caso el futuro docente, en el tema de estudio. Se dirá que el sujeto “comprende” si las prácticas que realiza al proponerle una tarea son adecuadas y consideradas correctas desde el punto de vista de la institución escolar. Siguiendo a estos autores, en el Capítulo 4 de la presente memoria se evalúa la comprensión sobre inferencia estadística que poseen los futuros profesores mediante el análisis de las respuestas (entendidas como prácticas), que elaboran al responder un conjunto de tareas que caracterizan el instrumento de evaluación propuesto para su evaluación. Dado que la comprensión personal sobre cierto objeto matemático, como la inferencia estadística, se debe a que la comprensión no se puede observar directamente entonces este tipo de evaluación se debe deducir de las respuestas en las pruebas de evaluación que tiene el estudiante, es decir, de sus prácticas personales.

Al considerar la actividad de evaluación, Godino (1996) sugiere que es necesario tener en cuenta dos ejes:

- *Eje descriptivo*, que indica los aspectos o componentes de los objetos que se pretenden evaluar, en nuestro caso, los contenidos de inferencia estadística que se incluyen en las pruebas de acceso a la universidad.
- *Eje procesual*, que indica las fases o niveles necesarios para considerar que un estudiante alcanza la comprensión; en las pruebas de acceso a la universidad será el tipo de respuesta, considerada adecuada, en cada clase de problema.

Dentro del EOS, tal y como se ha comentado anteriormente, la evaluación de la

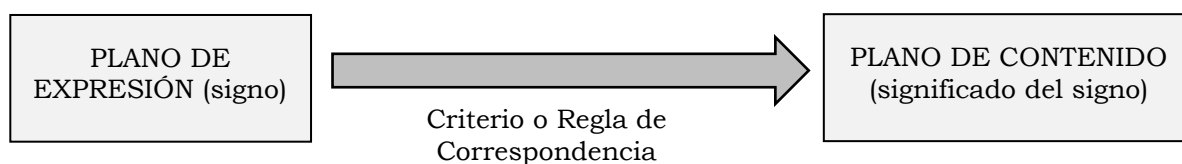
comprensión sería el estudio de la correspondencia entre los significados personales e institucionales. Por tanto, la comprensión depende de la institución desde donde se lleva a cabo la evaluación, que es la que determina hasta qué punto un sujeto “comprende” el significado de un objeto. En nuestro caso, son los responsables de establecer las pruebas los que deciden los criterios para asegurar la comprensión de la inferencia estadística por parte de los estudiantes.

1.5.4. Función semiótica y conflicto semiótico

El sistema de prácticas matemáticas que realiza una persona al resolver un problema pone en juego una gran diversidad de objetos matemáticos, de los diversos tipos descritos en la Subsección 1.5.1. Cuando un estudiante se enfrenta a un problema necesita interpretar expresiones matemáticas, en las que los símbolos remiten a conceptos, propiedades o procedimientos. Pero puede malinterpretar alguna de estas expresiones e igualmente puede utilizar una representación inadecuada cuando expone los resultados de su trabajo.

Para tener en cuenta esta realidad, el EOS adopta de otros autores la idea de función semiótica, entendida como una entidad que describe una relación entre diversos objetos matemáticos (Godino, 2002; Godino et al., 2007; 2019). Metafóricamente, la noción de función semiótica puede ser concebida como una correspondencia entre un antecedente (expresión, significantes) y un consecuente (contenido o significado), establecido por un sujeto (persona o institución) de acuerdo con un cierto criterio o regla de correspondencia (véase Figura 1.5.3). Además, según los autores, cualquiera de los tipos de objetos (problemas, conceptos, propiedades, lenguajes, procedimientos y argumentos) pueden ser expresión o contenido de una función semiótica.

Figura 1.5.3. *Función Semiótica*



Este tipo de relaciones no solo supone una correspondencia entre el antecedente y el consecuente, sino que permite proponer una interpretación del conocimiento y la comprensión de un objeto O por parte de un sujeto X (persona o institución) e implica generación de conocimiento por parte del agente interpretante (Godino, 2003). De este modo el conocimiento se concibe como el conjunto de funciones semióticas que una persona o institución es capaz de establecer a propósito de un cierto objeto O.

Por otro lado, el estudiante en su trabajo lleva a cabo sus propias funciones semióticas, que serán correctas si concuerdan con las admitidas dentro de la institución de enseñanza (Godino y Batanero, 1994). Cuando el interpretante no establece dicha correspondencia o cuando interpreta erróneamente una función semiótica propuesta por la institución, entonces se produce un *conflicto semiótico*, entendido como “la disparidad o desajuste entre los significados atribuidos a una misma expresión por dos sujetos (personas o instituciones) en interacción comunicativa” (Godino, 2002, p.14).

Burgos y Godino (2019) diferencian entre conflictos semióticos epistémicos cuando la disparidad o desajuste se produce entre significados de tipo institucional y conflictos semióticos cognitivos si la disparidad tiene lugar entre el significado manifestado por un sujeto y el de referencia. Uno de los fines del segundo estudio de evaluación llevado a cabo en esta tesis será identificar los posibles conflictos semióticos cognitivos de los profesores participantes en relación a la inferencia.

1.5.5. Idoneidad didáctica y sus componentes

Otro elemento del marco teórico que usaremos para valorar nuestras conclusiones es el de idoneidad didáctica, que Godino, Contreras y Font (2006) definen como un indicador de la calidad de un proceso educativo (o bien un material o un instrumento de evaluación, en nuestro caso), que se compone de seis componentes (Godino, 2011; 2013; Godino et al., 2013; Godino et al., 2006).

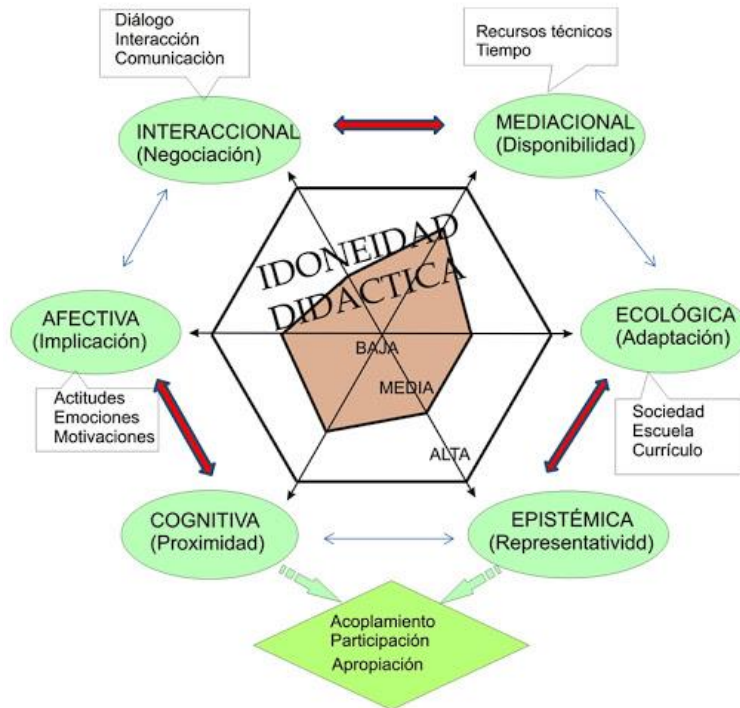
- *Idoneidad epistémica*, Indica el grado de concordancia y representatividad entre los significados institucionales implementados (pretendidos o evaluados), respecto de un significado de referencia. En nuestro caso, se evalúa la concordancia y representatividad entre el significado de la inferencia

estadística pretendido en el currículo y el significado evaluado en las pruebas de acceso a la universidad.

- *Idoneidad cognitiva*, grado en el que los significados pretendidos/implementados son asequibles a los estudiantes. En el estudio que estamos realizando, no los podremos deducir directamente, sino por la información de otros trabajos realizados con alumnos, o indirectamente, mediante el análisis de la dificultad potencial de los problemas.
- *Idoneidad interaccional*, grado en el que las configuraciones y trayectorias didácticas organizadas en el aula permiten identificar conflictos semióticos potenciales y resolver conflictos observados durante el proceso de instrucción. Este componente no se aplica en nuestro estudio.
- *Idoneidad mediacional*, grado de disponibilidad y adecuación de los recursos materiales y temporales necesarios para el desarrollo del proceso de enseñanza-aprendizaje. En estas pruebas el principal material sería el tiempo disponible, así como las calculadoras y las tablas de probabilidades de la distribución normal que se proporciona al estudiante.
- *Idoneidad emocional*, grado de implicación (por ejemplo, interés o motivación) del alumnado en el proceso de estudio. Lo valoraremos analizando los contextos de los problemas y su posible interés para el estudiante.
- *Idoneidad ecológica*, grado en que el proceso de estudio se ajusta al proyecto educativo de la institución (o comunidad) y a los condicionamientos del entorno en que se desarrolla (Godino et al., 2006).

La Figura 1.5.4 corresponde al esquema de idoneidad didáctica descrito por Godino (2013) en el que además de identificar las seis componentes, el autor sugiere los criterios que permiten evaluar la idoneidad didáctica. Su finalidad es determinar cualitativamente el grado de idoneidad didáctica del proceso de enseñanza-aprendizaje permitiendo su mejora. La incorporación de estos criterios permite orientar sobre cómo realizar el proceso de enseñanza-aprendizaje y valorar la efectividad de este una vez implementado.

Figura 1.5.4. Idoneidad Didáctica



Nota: Tomado de Godino (2011, p. 6).

1.5.6. Modelo del conocimiento del profesor en el EOS

La necesidad de formación de profesores para enseñar estadística es un problema internacional, que llevó a la *International Association for Statistical Education* y a la *International Committee for Mathematical Instruction* a organizar un estudio conjunto para discutir esta problemática (Batanero, Burrill y Reading, 2011). Dicho estudio ha impulsado notablemente la investigación sobre formación de profesores de estadística.

Anteriormente encontramos autores que inciden en la necesidad de analizar dicha formación (Batanero, Godino y Roa, 2004; Godino et al., 2008; Mendonça, Coutinho y Almouloud, 2006; Stohl, 2005). Batanero (2001) indicó que pocos futuros profesores de secundaria reciben una formación específica en estadística aplicada o didáctica de la estadística, a la que no pueden transferirse algunos principios generales válidos para otras ramas de las matemáticas.

Por tal motivo, es necesario recurrir a entornos y contextos significativos en los

programas de formación de futuros profesores que fomenten procesos de reflexión sobre dichas actividades (Llinares y Krainer, 2006). Este tema ha generado un gran número de investigaciones por el impacto que tiene dentro del ámbito de la Educación Estadística, algunas de las cuales se recogen en Batanero et al. (2011). En el seno del propio grupo de investigación se han elaborado algunas tesis doctorales centradas en los conocimientos de los profesores para enseñar temas específicos de estadística o matemáticas, como las de Arteaga (2011), Burgos (2020), Contreras (2011) o Gea (2014), entre otras.

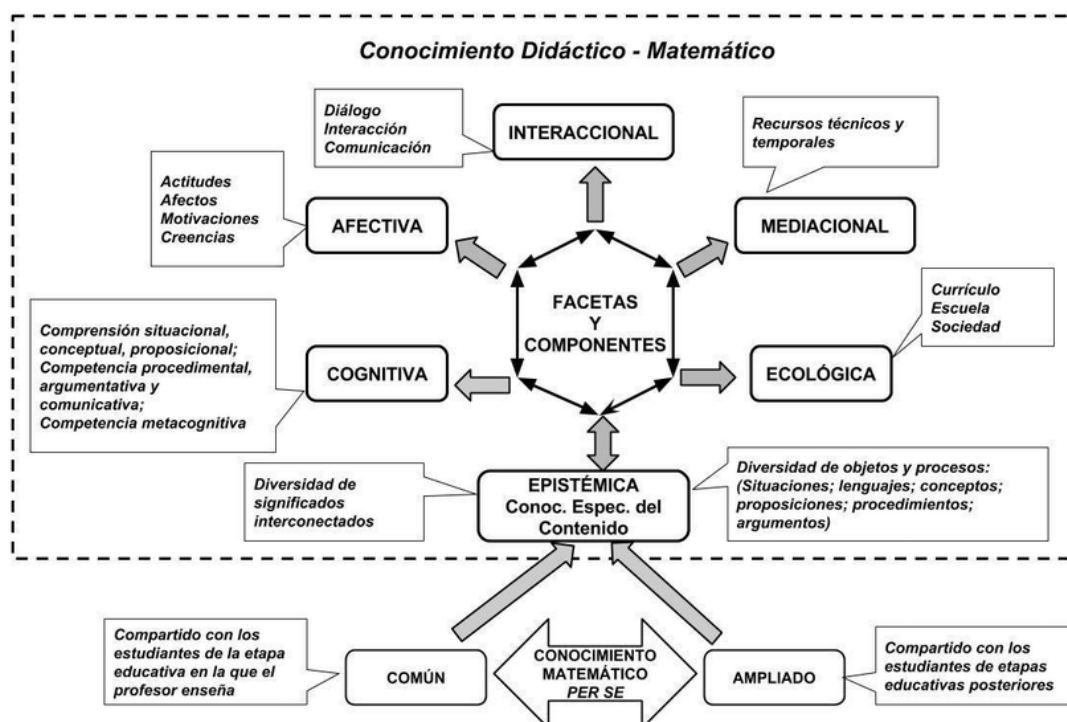
Son muchos los modelos sugeridos en Didáctica de la Matemática para describir los conocimientos que un profesor requiere para poder enseñar un tema de matemáticas con éxito y resaltar la complejidad de la labor del profesor. Ejemplos notables son los de Shulman (1986; 1987) y los de Ball y colaboradores (Ball et al., 2008; Hill et al., 2008). Estos modelos se han aplicado a diferentes investigaciones y publicado en revistas como *Journal of Mathematics Teacher Education*. Otros se han adaptado al campo de la educación estadística, véase, por ejemplo, la descripción de algunos de estos modelos en Godino et al. (2011) o el trabajo de Groth (2007).

Godino (2009) señala que los modelos elaborados desde las investigaciones en Educación Matemática incluyen categorías muy generales y destaca la importancia de contar con modelos que permitan un análisis más detallado que permita orientar el diseño de acciones formativas y la elaboración de instrumentos de evaluación de los conocimientos del profesor de matemáticas. Motivado por este tema, Godino y colaboradores (Godino, 2009; 2011; Godino et al., 2017; Pino-Fan y Godino, 2015) proponen el modelo del conocimiento didáctico-matemático del profesor (CDM) organizado en la dimensión matemática y la didáctica, cada una con varios componentes (Figura 1.5.5). Tomaremos este modelo en la investigación descrita en esta memoria.

El modelo CDM asume que el profesor debe tener un conocimiento matemático común relativo al conocimiento del contenido pretendido en el nivel educativo donde imparte su docencia, y otro conocimiento matemático ampliado que le permita articularlo con los niveles superiores de enseñanza. Para el Estudio 2 descrito en el

Capítulo 4 nos interesaremos por evaluar el conocimiento común de los futuros profesores de la muestra respecto al contraste de hipótesis y el intervalo de confianza.

Figura 1.5.5. *Facetas y Componentes del Conocimiento Didáctico-Matemático*



Nota: Tomado de Godino et al. (2016, p. 289).

Asimismo, los autores señalan que el profesor deber tener un conocimiento didáctico-matemático, que abarca las distintas facetas que afectan en el proceso educativo necesarias para la enseñanza y aprendizaje de contenidos específicos, relacionadas con las ya descritas de la idoneidad didáctica (véase Figura 1.5.6). Las seis facetas incorporadas en el modelo CDM son:

- *Faceta epistémica*, es el conocimiento didáctico sobre el propio contenido y está relacionada con la idoneidad epistémica y supone el conocimiento de los significados institucionales del tema que se pretende enseñar, incluyendo los objetos matemáticos que los caracterizan (problemas, lenguaje, conceptos, propiedades, argumentos y procedimientos) y procesos relacionados con ellos.
- *Faceta cognitiva*, implica el conocimiento de la forma en que los estudiantes

aprenden, razonan y resuelven problemas del tema dado, las fases de su desarrollo cognitivo, sus estrategias de aprendizaje y de resolución de problema, sus posibles errores y obstáculos, entre otros. Evaluaremos también algunos elementos de esta faceta en el Estudio 2, preguntando a los futuros profesores qué tipos de errores en inferencia prevén en sus futuros estudiantes.

- *Faceta afectiva*, conocimiento del grado de implicación (interés o motivación) del alumnado en el proceso de estudio, de sus sentimientos y de toda la componente emocional (actitudes, emociones, creencias).
- *Faceta mediacional*, conocimiento de recursos didácticos disponibles para la enseñanza y como usarlos en el tema en particular, así de cómo organizar el trabajo con los mismos según el nivel de formación o el grado al cual se imparte la enseñanza. Para evaluar algún elemento de esta faceta, se preguntará a los futuros profesores en el Estudio 2 la forma en que enseñarían a sus estudiantes el intervalo de confianza.
- *Faceta interaccional*, conocimiento de modelos de comunicación entre estudiantes y profesor para que sea posible hacer aflorar los conflictos semióticos de los estudiantes y que se produzca el aprendizaje.
- *Faceta ecológica*, conocimiento del grado en que el proceso de estudio se ajusta al currículo, al proyecto educativo de la institución), relaciones con otras materias y la vida cotidiana y laboral.

Además, para cada una de las anteriores facetas, el autor propone promover la capacidad de análisis didáctico del profesor en cuatro niveles diferentes como muestra la Figura 1.5.6 (Godino et al., 2011).

1. *Análisis de prácticas matemáticas y didácticas*. Consiste en desarrollar la capacidad del profesor para identificar problemas adecuados para un tema particular; analizar las prácticas operativas y discursivas necesarias (a priori) o realizadas (a posteriori) para resolver estas tareas. También la capacidad de identificar las líneas generales de actuación del docente y discentes con el fin de conseguir un aprendizaje particular y contextualizar el contenido.
2. *Descripción de las configuraciones de objetos y procesos matemáticos*. Se trata

de identificar los objetos y procesos probabilísticos que intervienen o emergen en la realización de las prácticas anteriormente citadas; con énfasis en concienciarse de la complejidad, de contenido y su didáctica, y valorar sus posibles significados, diferenciados como factor explicativo de los conflictos y de la progresión del aprendizaje.

3. *Identificación del sistema de normas y metanormas.* Se estudian los fenómenos de índole social a través de una dimensión normativa, la cual condiciona y hace posible el proceso de enseñanza y aprendizaje de la probabilidad.
4. *Valoración de la idoneidad didáctica del proceso de estudio.* Sintetiza los análisis previos con el fin de valorar el proceso de estudio como un todo e identificar potenciales mejoras para futuras implementaciones.

Figura 1.5.6. *Facetas y Niveles del Conocimiento del Profesor del Modelo CDM*



Nota: Tomado de Godino (2009, p. 21).

1.6. OBJETIVOS E HIPÓTESIS DE LA INVESTIGACIÓN

Finalizada la exposición del tema y la justificación de su importancia, a continuación, se describen los objetivos que se persiguen con la presentación de la memoria y se formulan algunas hipótesis sobre lo que se espera encontrar.

1.6.1. Objetivos del trabajo

Tal y como ha sido señalado en la introducción, la investigación que se desea desarrollar se orienta en analizar la presencia de la inferencia estadística en las pruebas de acceso a la universidad y algunos componentes del conocimiento didáctico-matemático de futuros profesores de Educación Secundaria Obligatoria y Bachillerato sobre inferencia estadística. Todo ello surge con la finalidad de proponer acciones formativas para mejorar la formación tanto de los estudiantes, como de los futuros profesores. El objetivo principal del trabajo puede desglosarse en los siguientes objetivos específicos:

OE1. Realizar una síntesis de la investigación sobre dificultades en la comprensión y aplicación de la inferencia estadística y del análisis de los problemas propuestos en las pruebas de acceso a la universidad. Previo a los estudios empíricos propuestos, es importante analizar las investigaciones desarrolladas en torno al tema de estudio, para fijar nuestro trabajo en un contexto más amplio.

El interés de este objetivo radica, en primer lugar, en que el conocimiento sobre los errores más frecuentes o dificultades que tienen los estudiantes cuando trabajan los contenidos de inferencia estadística, supone una ayuda a los docentes. El hecho de contar con esta información permite a los profesores adaptar su proceso de enseñanza-aprendizaje con el fin de que los estudiantes asimilen, de una forma correcta, los contenidos incluidos en el bloque objeto de estudio. Para lograr este objetivo en el Capítulo 2 se realiza una síntesis de investigaciones previas localizando, analizando y clasificando un gran número de investigaciones en relación con diferentes objetos matemáticos que forman parte de la inferencia estadística.

Igualmente se sintetiza la investigación sobre la inferencia estadística en los libros de texto, el análisis de las pruebas de acceso a la universidad y de otras pruebas de evaluación. Esta parte de la síntesis de antecedentes permite determinar qué aspectos de las pruebas de acceso a la universidad han sido objeto de investigación y cuáles pueden ser completados.

OE2. Caracterizar los problemas de inferencia estadística propuestos en las pruebas de acceso a la universidad y comparar el significado evaluado en dichas pruebas con el significado institucional pretendido para la inferencia estadística en

las orientaciones curriculares.

El enfoque ontosemiótico (Godino et al., 2007) asume que los objetos matemáticos surgen de las prácticas matemáticas realizadas para resolver problemas relacionados con el objeto, considerando el término problema en sentido amplio. Por ello, es importante conocer con el mayor detalle posible las características de los problemas propuestos en las pruebas de evaluación. Para lograr este objetivo, y en base al enfoque ontosemiótico, se ha resuelto cada uno de los problemas de las pruebas de acceso a la universidad de la comunidad autónoma andaluza para la materia Matemáticas Aplicadas a las Ciencias Sociales II durante el periodo 2003-2014. La resolución permitirá identificar la configuración de objetos matemáticos relacionados con la inferencia estadística requeridos en su solución y determinar el significado institucional evaluado en las pruebas. En el Capítulo 3 se describe con detalle el estudio llevado a cabo y sus conclusiones.

El interés de este estudio es que los problemas propuestos a los estudiantes, bien durante la enseñanza, o en las pruebas de evaluación, van a incidir directamente en el aprendizaje de diferentes objetos matemáticos y el significado asignado a los mismos. La razón es que los profesores tratan de preparar a los estudiantes para superarlas con la mejor nota posible, teniendo también en cuenta que la nota en estas pruebas concretas pueden determinar la carrera que el alumno finalmente curse.

OE3. Evaluar algunos elementos del conocimiento didáctico-matemático sobre la inferencia estadística en futuros profesores de Educación Secundaria Obligatoria y Bachillerato. Más concretamente, se desea evaluar algunos elementos de su conocimiento común del contenido y de la faceta cognitiva de su conocimiento didáctico.

La investigación sobre conocimientos de inferencia estadística de los futuros profesores es prácticamente inexistente y por ello hemos considerado de interés llevar a cabo este estudio, incluso con carácter parcial y exploratorio. Por ello, basándose en las investigaciones anteriores, se prepara un cuestionario siguiendo los pasos descritos en la metodología, asegurando la calidad del instrumento (McDonald, 2013). La necesidad del instrumento se justifica por no haber encontrado en los antecedentes uno que sirva para nuestros objetivos. La aplicación del cuestionario y el análisis cualitativo detallado

de las respuestas al mismo nos ha permitido evaluar la componente epistémica del conocimiento didáctico-matemático que tienen los futuros profesores de Educación Secundaria Obligatoria y Bachillerato, y por otro, proponer acciones formativas para mejorarlo. Los resultados de la evaluación se describen en el Capítulo 4.

1.6.2. Hipótesis o expectativas iniciales

Una vez expuestos los objetivos, se formulan algunas hipótesis, que se entiende cómo expectativas iniciales y que se discutirán en las conclusiones en el Capítulo 5. Señalar que no realizaremos contrastes de hipótesis en el sentido formal de la inferencia estadística. Estas hipótesis son las siguientes:

H1. El análisis de las investigaciones previas mostrará que el tema que nos ocupa ha sido escasamente investigado y, por tanto, nuestro trabajo puede aportar información original.

Para formular esta hipótesis nos basamos en nuestra primera búsqueda bibliográfica, realizada durante la elaboración del Estudio 1, en la que sólo tres trabajos están ligeramente relacionados con el actual. Igualmente nos apoyamos en el estudio de los antecedentes realizado por Carretero (2014), que tampoco encontró investigaciones similares a la que ahora abordamos. Una explicación a esta escasez de investigaciones es la dificultad del tema; se requiere un conocimiento estadístico profundo para abordar una investigación didáctica sobre inferencia estadística. Pero la mayoría de los investigadores en Educación Matemática no son estadísticos, ni suelen ser profesores de inferencia estadística. Es por ello por lo que el tema ha recibido poco interés dentro de la Educación Matemática.

Por otro lado, al analizar algunos trabajos de síntesis sobre inferencia estadística (por ejemplo, Batanero, 2020 o Harradine et al., 2011) y las tesis doctorales sobre inferencia estadística realizadas en nuestro grupo de investigación (Alvarado, 2007; Olivo, 2008; Vallecillos, 1994) se constató que casi todas las investigaciones sobre comprensión de la inferencia estadística han sido llevadas a cabo con estudiantes universitarios de diferentes especialidades. Por ello, se decidió realizar una búsqueda bibliográfica más amplia para analizar los aspectos no tratados en la investigación sobre

conocimiento inferencial de los futuros profesores.

H2. El análisis semiótico de los problemas de inferencia propuestos en las pruebas de acceso a la universidad y su comparación con el contenido curricular de estadística en los decretos curriculares de Bachillerato mostrará, por un lado, que se da una alta importancia a la inferencia en las pruebas de acceso. Además, se observará un buen ajuste entre los contenidos de inferencia fijados en el currículo y los considerados en dichas pruebas.

Nos basamos para formular esta hipótesis en un análisis preliminar publicado por Díaz, Mier, Alonso y Rodríguez-Muñiz (2014). Otra serie de hipótesis sobre estos problemas se exponen y discuten en el Capítulo 3.

H3. Los estudios de evaluación llevados a cabo con futuros profesores mostrarán en general una buena preparación matemática sobre el tema. Sin embargo, esperamos encontrar algunos conflictos semióticos cognitivos relacionados con el tema y sobre todo de interpretación de los conceptos implicados en la inferencia estadística. Además, esperamos mostrar un conocimiento didáctico insuficiente de dichos profesores en las facetas evaluadas en nuestro estudio.

Parte de esta hipótesis se basa en los resultados sobre dificultades de comprensión de la inferencia realizados con alumnos universitarios de licenciaturas similares a las que podrían tener los futuros profesores de la nuestra. Pensamos que los errores encontrados en otros tipos de estudiantes universitarios se encontrarán en los estudiantes que se preparan para ser profesores de matemáticas en la etapa educativa de secundaria y bachillerato. La parte final la conjeturamos del hecho que estos futuros profesores no reciben en la actualidad ningún tipo de formación didáctica sobre inferencia estadística.

1.7. ORGANIZACIÓN DE LA INVESTIGACIÓN

Para finalizar este capítulo se hace un breve resumen de las principales características de la investigación y la forma en que ha sido organizada. El enfoque general de la investigación es descriptivo y exploratorio, con algunos elementos interpretativos y explicativos, ya que tratamos de interpretar los datos obtenidos a la luz del marco

teórico. No se realiza un control y manipulación de variables independientes, sino que se engloba en la investigación cuasi-experimental (Shadish et al., 2002).

De acuerdo a Bisquerra y Alzina (2004) podemos clasificar nuestro trabajo como investigación aplicada, puesto que sus resultados pueden utilizarse para la mejora de la enseñanza de la inferencia estadística. Por sus fuentes, es empírica, pues se basa en la recogida y análisis de datos. Para Tamayo (2004), el estudio con profesores se trataría de un diseño transversal, puesto que tomamos los datos en una sola ocasión. El estudio de las pruebas de acceso a la universidad sería un diseño ortogonal o equilibrado pues el número de problemas analizados cada año es similar.

Una vez que se han descrito aquellos puntos que permiten enmarcar y caracterizar la investigación que llevaremos a cabo en la tesis doctoral, en este apartado describimos la organización de la memoria que describe los resultados de esta, y que consta de dos partes bien diferenciadas. La primera parte de la memoria hace referencia a los fundamentos del trabajo, mientras que la segunda parte presenta los resultados de los dos estudios empíricos descritos (véase Figura 1.7.1).

Los fundamentos utilizados para enmarcar el trabajo son los siguientes:

- *Análisis curricular*, que se ha presentado en el Capítulo 1, y que engloba el estudio de dos directrices curriculares: en primer lugar, el currículo oficial estatal (MEC, 2007), por el que se establece las enseñanzas mínimas correspondientes a la etapa de Educación Secundaria Obligatoria y en segundo la Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE) que se recoge en el Currículo Básico (MECD, 2015b). El análisis de las normativas curriculares permitirá, en una primera fase, enmarcar el significado institucional pretendido que tiene la inferencia estadística en Bachillerato, para posteriormente, realizar su comparación con el significado institucional recogido en las pruebas de acceso a la universidad.
- *Marco teórico*, que fundamente la investigación y que es el enfoque ontosemiótico de la cognición e instrucción matemática (Godino, 2002; 2003; Godino y Batanero, 1994; Godino et al., 2007; 2020). Este marco teórico se ha venido proponiendo y utilizando a lo largo de más de veinte años dentro de nuestro equipo de investigación y en otros grupos de investigación. Dicho

marco teórico se ha presentado en este capítulo, donde se introduce los elementos del enfoque ontosemiótico que se utilizarán en la parte empírica. En particular nos centramos en las ideas de objeto matemático, prácticas y significados, y sus configuraciones epistémicas; comprensión y evaluación; función semiótica y conflicto semiótico; idoneidad didáctica y sus componentes y por último en algunos componentes del modelo del conocimiento didáctico-matemático del profesor.

- *Antecedentes de investigación*, es decir los trabajos previos que permiten ubicar nuestro trabajo dentro de una línea más amplia de investigación y discutir sus resultados. Serán también utilizados para comparar con nuestros resultados validando las conclusiones obtenidas en nuestro trabajo. La descripción de las investigaciones relacionadas con nuestro objeto de estudio se presenta con detalle en el Capítulo 2, organizado en torno a dos grupos principales de trabajos: comprensión de conceptos de inferencia estadística y análisis de pruebas de evaluación.

Una segunda parte de la memoria consiste en la descripción de dos estudios empíricos. A continuación, se resumen las características metodológicas generales empleadas en los dos estudios realizados, que se describen con detalladamente en el Capítulo 3 y en el Capítulo 4.

- El *primer estudio*, de carácter empírico (Estudio 1), trata de un estudio documental de los problemas de inferencia propuestos a lo largo de 12 años en las pruebas de acceso a la universidad en Andalucía (en total 144 problemas).

Estos problemas se han clasificado en campos de problemas, y para cada uno de ellos se ha resuelto un ejemplo característico, identificando las prácticas necesarias en su resolución y la configuración epistémica implicada. Posteriormente se llevó a cabo un análisis de contenido del enunciado y su solución de los 144 problemas (Krippendorf, 2013). Se han identificado las variables que los caracterizan, y se realiza un estudio estadístico descriptivo de las categorías de las mismas y su evolución temporal.

El Capítulo 3 recoge el análisis detallado de los problemas sobre inferencia estadística permitiéndonos analizar la idoneidad didáctica de las mencionadas pruebas.

- El *segundo estudio*, de carácter también empírico (Estudio 2) consiste en un estudio de evaluación de los conocimientos didáctico- matemáticos que muestra un grupo de futuros profesores de Educación Secundaria y Bachillerato sobre contenidos de inferencia estadística utilizando un cuestionario de elaboración propia.

Figura 1.7.1. Organización del trabajo



Concretamente, se evalúa su conocimiento común del contraste de hipótesis e intervalo de confianza y algunos elementos de las facetas cognitivas y mediacional de dicho conocimiento. La metodología es cualitativa, basada en el análisis de las respuestas escritas de los participantes a las tareas propuestas. La información recogida en el Capítulo 4 permite obtener conclusiones sobre el conocimiento matemático que poseen las personas responsables de la enseñanza de la inferencia estadística en los centros de secundaria y bachillerato. Asimismo, el estudio realizado en el mencionado capítulo se completa con el análisis de los conocimientos didáctico-matemático que poseen los mencionados participantes.

ANTECEDENTES DE LA INVESTIGACIÓN

Índice de Contenidos

- 2.1. Introducción
- 2.2. Investigaciones sobre comprensión del contraste de hipótesis
 - 2.2.1. Críticas al uso del contraste de hipótesis
 - 2.2.2. Comprensión del concepto de hipótesis
 - 2.2.3. Comprensión del nivel de significación y el p -valor
 - 2.2.4. Comprensión del proceso de contraste
- 2.3. Comprensión del intervalo de confianza
 - 2.3.1. Definición del intervalo y del nivel de confianza
 - 2.3.2. Propiedades del intervalo de confianza
- 2.4. Dificultades con otros conceptos de inferencia
 - 2.4.1. Comprensión de la distribución muestral
 - 2.4.2. Comprensión del teorema central del límite
- 2.5. Análisis de problemas de inferencia estadística
 - 2.5.1. Análisis de la inferencia estadística en libros de texto
 - 2.5.2. Análisis de las pruebas de acceso a la universidad
 - 2.5.2.1. Análisis de las pruebas de matemáticas
 - 2.5.2.2. Estudio de la inferencia estadística en las pruebas de acceso a la universidad
 - 2.5.3. Análisis de otras pruebas de evaluación
- 2.6. Conclusiones de los antecedentes
 - 2.6.1. Conflictos identificados en la literatura
 - 2.6.2. Posibles causas de los conflictos descritos
 - 2.6.3. Conclusiones sobre las investigaciones relacionadas con las pruebas de acceso a la universidad

2.1. INTRODUCCIÓN

Una vez mostrados los objetivos que se persiguen con el desarrollo de esta memoria y las hipótesis iniciales sobre los resultados que esperamos obtener en la investigación, se presenta en este capítulo un resumen de las investigaciones previas que sustentan el trabajo.

Dentro del proceso de enseñanza y aprendizaje de las matemáticas un elemento a tener en cuenta, a la hora de tratar cualquier tema, son las posibles dificultades que pueden tener los estudiantes. El estudio o análisis de dichas dificultades, resumido en los siguientes apartados para el caso particular de la inferencia estadística, puede también contribuir a la mejora de la enseñanza y aprendizaje de los estudiantes. Además, es previsible que las dificultades descritas en la literatura previa en estudiantes o en profesionales que aplican la inferencia estadística sean compartidas por los futuros profesores.

La inferencia es el tema de estadística en que probablemente hay un mayor número de estudios sobre errores de comprensión, debido a que este tema ha interesado no sólo a investigadores en didáctica, sino a estadísticos y psicólogos. Una enseñanza desde un enfoque formal deductivo y basada en nociones tan abstractas, como, por ejemplo, la distribución muestral, el intervalo de confianza, los estadísticos de contraste, la definición de hipótesis nula y alternativa, el nivel de confianza y nivel de significación, ocasionan grandes dificultades para los estudiantes (Harradine et al., 2011). Con frecuencia, además, la inferencia es enseñada a estudiantes con poca base de álgebra y cálculo y, por tanto, se consigue una comprensión muy superficial del tema (Artigue et al., 2007).

A continuación, se describen las investigaciones centradas en la descripción de los errores más comunes en la comprensión o aplicación de la inferencia estadística. De las investigaciones desarrolladas en este campo, se han seleccionado las centradas en alumnado de bachillerato o alumnado universitario. Las primeras, porque las pruebas de acceso a la universidad están dirigidas a estos estudiantes y las segundas, porque, hay pocos antecedentes sobre los conocimientos de los profesores y futuros profesores que han recibido formación sobre este tema en su etapa universitaria. Una información

más amplia sobre estas dificultades se puede encontrar en Batanero y Borovcnik (2016), Batanero y Díaz-Batanero (2006), Batanero et al. (2020), Castro-Sotos et al. (2007), Harradine et al. (2011), Tobias-Lara y Gomez-Blancarte (2019) y Vallecillos (1994, 1999).

Se inicia la revisión de los errores descritos en inferencia estadística con aquellos relacionados con el contraste de hipótesis, cuyo uso en la investigación empírica ha sido criticado desde su creación, incluso por científicos tan prominentes como Yates (Batanero, 2000; Díaz-Batanero, Lozano-Rojas y Fernández-Calderón, 2019). Por ello se incluye una sección describiendo esta controversia y las recomendaciones actuales para mejorar la práctica de la estadística.

Dada la relevancia señalada por diversos organismos en relación al intervalo de confianza como alternativa al uso de la significación estadística (Cumming y Finch, 2005; Cumming et al., 2012), una segunda sección de esta revisión se centra en el análisis de los errores en la comprensión de diversos objetos que intervienen en el intervalo de confianza.

Finalmente se describen las investigaciones relacionadas con la comprensión de conceptos ligados a ambos bloques, como son la distribución muestral y el teorema central del límite.

Puesto que parte de nuestro trabajo se centra en el análisis de los problemas de inferencia estadística planteados en las pruebas de acceso a la universidad en Andalucía, asimismo, se integran aquellas investigaciones centradas en el análisis del tema de inferencia estadística en los libros de texto y en las pruebas de acceso. Dado que el tema en cuestión está recogido en el área de matemáticas, se presenta, por un lado, los antecedentes de investigaciones relacionadas con análisis de las pruebas de matemáticas y se prosigue con aquellos trabajos que han centrado su atención en el estudio de la inferencia estadística en las pruebas de acceso a la universidad. Seguidamente, y en la misma línea que el punto anterior, se analizan algunas de las investigaciones llevadas a cabo sobre otras pruebas de evaluación.

Para finalizar, se recogen las conclusiones obtenidas de esta síntesis de la literatura, que puede ser ampliada en Batanero, Díaz-Batanero y López-Martín (2017); Batanero et al., (2020) y López-Martín, Molina-Portillo, Contreras y Ruz, (2019).

2.2. INVESTIGACIONES SOBRE COMPRENSIÓN DEL CONTRASTE DE HIPÓTESIS

La realización de un contraste de hipótesis requiere el manejo y comprensión de un conjunto de conceptos que, con a pesar de tener apariencia similar, tienen importantes diferencias (Inzunza y Jiménez, 2013; Vallecillos, 1994). Así, por ejemplo, tenemos los conceptos de población y muestra, hipótesis nula y alternativa, error tipo I y II, región crítica (o de rechazo) y de aceptación, estadístico y parámetro (Batanero, 2000).

Por ello, dentro de la investigación sobre didáctica de la estadística, una gran variedad de trabajos ha descrito errores e interpretaciones incorrectas relacionadas con la comprensión del contraste de hipótesis (Batanero, 2000; Falk y Greenbaum, 1995). Algunas de estas dificultades tienen su origen, en parte, en una deficiente comprensión del concepto de probabilidad condicional, pues algunos conceptos de inferencia estadística, como los de nivel de significación y p -valor, así como la distribución muestral del estadístico o la potencia se definen usándola (Batanero y Borovcnick, 2016; Falk, 1986).

Hasta nuestro conocimiento, son muy pocos los trabajos realizados con estudiantes de bachillerato. Destacamos el llevado a cabo por Espinel et al. (2007) donde las autoras realizan una lista de los errores más frecuentes que cometen los estudiantes cuando abordan las cuestiones de inferencia estadística incluidas en las pruebas de acceso a la Universidad de La Laguna. Los errores que detectan son: 1) dificultad al entender la noción de probabilidad; 2) tipificación errónea de la distribución Normal; 3) uso incorrecto de la tabla de probabilidades de la Normal; 4) fallo en la aproximación de la Binomial a la Normal; 5) confusión entre muestra y población; 6) descuido ante cambios del tamaño muestral; 7) dificultad para discernir entre parámetro, estimador y estimación; 8) identificación y formulación confusa de los contrastes unilaterales y bilaterales; 9) intercambio del valor empírico con el valor muestral; 10) asignación incorrecta del nivel de confianza; 11) asociación errónea del nivel de significación y la probabilidad buscada; 12) interpretación errónea de las conclusiones en los contrastes. Estos errores son analizados con detalle en Ramos et al. (2009).

La aplicación incorrecta del contraste de hipótesis también detectado en la investigación empírica ha suscitado un debate en relación a la aplicación de método

estadístico en investigación. En lo que sigue se analiza con detalle cada uno de estos aspectos, comenzando por un resumen de la controversia en torno al contraste de hipótesis y seguidamente por la investigación realizada con estudiantes universitarios, profesionales y futuros profesores.

2.2.1. Críticas al uso del contraste de hipótesis

A lo largo de la historia reciente, las ciencias empíricas han apoyado la obtención de nuevo conocimiento a través de la observación de la realidad. Este principio, base del razonamiento inductivo, hace difícil la validación de teorías científicas. La inferencia estadística, con sus distintas tradiciones, ha tratado de dar solución al problema de justificación del razonamiento inductivo (Hacking, 2006).

Por su parte, la aproximación frecuencial, que es la enseñada en bachillerato y en la que basamos nuestro trabajo, utiliza un razonamiento hipotético-deductivo. Por un lado, a partir de la evidencia obtenida en muestras aleatorias se estima el valor de un parámetro, pero el cálculo de un p -valor se obtiene tomando como verdadera una hipótesis nula de partida (Banegas et al., 2000).

En contraste, la aproximación bayesiana, plantea el cálculo de la probabilidad de que una hipótesis sea cierta. Así, basada en la probabilidad subjetiva y el uso intenso del teorema de Bayes, estas metodologías difieren no sólo en la conceptualización de la probabilidad que aplican, sino en otros conceptos. Por ejemplo, el concepto de distribución muestral, aplicado en inferencia frecuencial, no es utilizado en inferencia bayesiana (Batanero, 2000; Díaz-Batanero, 2018; Díaz-Batanero et al., 2019).

Como se ha mencionado, estos dos tipos de inferencia tienen también objetivos y filosofías diferentes. Sin embargo, en la práctica los investigadores las mezclan, aplicando un procedimiento frecuencial al que dan una interpretación Bayesiana. Gigerenzer (1993) denomina *lógica híbrida* a este modo de aplicación de la inferencia estadística, que, según el autor, es rechazada tanto por los estadísticos frecuenciales como por los bayesianos. Éste y otros errores en la aplicación de la inferencia estadística han llevado a una gran crítica del método estadístico dentro de estas ciencias (Díaz-

Batanero et al., 2019; Gigerenzer, 1993; Harlow et al., 2016, Mayo y Cox, 2006; Valera et al., 2000).

La tendencia habitual durante años de aceptar artículos de investigación en algunas revistas únicamente si los contrastes estadísticos aplicados proporcionaban resultados estadísticamente significativos, denominado “sesgo de publicación” (De la Fuente y Díaz- Batanero, 2004; Harlow, Mulaik y Steiger, 2016; Valera et al., 2000) ha podido contribuir a la proliferación de algunos de estos errores. Por ejemplo, el incremento de la probabilidad de error tipo I al realizar numerosos contrastes sobre el mismo conjunto de datos (Hunter, 1997) o el uso arbitrario de niveles de significación 0,05 o 0,01 sin justificación, que Gigerenzer (2003) señala forma parte del ritual estadístico.

Un resultado de este debate fue el inicio por sociedades como la *American Psychological Association* de comités para mejorar la práctica estadística en las publicaciones del área (Wilkinson, 1999). Entre otras cosas, se recomendó complementar el contraste de hipótesis con la inclusión de intervalos de confianza y estimación del tamaño del efecto (APA, 2010). Sin embargo, Díaz- Batanero et al. (2019) indican que no se ha apreciado una mejora en la aplicación en inferencia en psicología y recomiendan una mejora en la enseñanza de la inferencia estadística para paliar el problema. Dicha mejora debería estar basada en el conocimiento de las dificultades que tienen en el tema estudiantes y profesores y en la mejor formación de estos últimos, no sólo en los aspectos matemáticos, sino en los didácticos. Este es el objetivo del estudio de evaluación realizado en el Capítulo 4 con futuros profesores y también de la síntesis de dificultades de los estudiantes que se realiza a continuación.

2.2.2. Comprensión del concepto de hipótesis

Un error común en los contrastes de hipótesis es el planteamiento de la hipótesis nula y alternativa, que, con frecuencia son confundidas por los estudiantes (Díaz-Batanero, 2018). Tal y como se establecen las hipótesis, algunos estudiantes consideran que el objetivo de dicho proceso consiste en demostrar la veracidad de la hipótesis nula, no considerando que es esta la que se quiere rechazar.

La creencia de que rechazar la hipótesis nula supone demostrar que es errónea, también se encontró en las investigaciones de Saldanha y Thompson (2002, 2007). Los autores señalan que tanto la idea de probabilidad y como la de atipicidad son fundamentales para entender la lógica que rige este tipo de pruebas, en las que se rechaza una hipótesis nula cuando una muestra de una población es considerada lo suficientemente atípica si la hipótesis nula fuese cierta.

En Cañadas et al. (2012), un 46% de estudiantes de psicología no logran formular correctamente las hipótesis. Manifiestan errores relacionados con la confusión entre: hipótesis nula y la alternativa, parámetro y estadístico, contraste de homogeneidad e independencia, valor y variable. Asimismo, los autores señalan errores asociados a la formulación de hipótesis que no cubren todo el espacio paramétrico.

Por su parte, Batanero et al. (2012) encontraron errores en el planteamiento de las hipótesis estableciendo contrastes de hipótesis unilaterales en lugar de contrastes de hipótesis bilaterales y viceversa. Del mismo modo los autores destacan errores vinculados con el uso de los estadísticos muestrales a la hora de determinar las hipótesis nula y alternativa, sin considerar los parámetros de la población. De igual modo que Cañadas et al. (2012), encuentran situaciones en las que se plantean hipótesis nulas y alternativas que no cubren el espacio paramétrico.

Por otro lado, Batanero (2000) señala dificultades en la distinción entre las hipótesis estadísticas y las hipótesis de investigación. Hay diferentes formas de concretar las hipótesis de investigación, que suelen ser muy generales. Por ejemplo, un investigador puede conjeturar que los estudiantes de un grupo experimental, donde se ha aplicado una enseñanza innovadora, han tenido mejor aprendizaje de las fracciones que los de un grupo control. Esta sería una hipótesis de investigación, mientras que una hipótesis estadística asociada sería posiblemente que la puntuación media en un cuestionario sobre las fracciones aplicado en los dos grupos es mayor en el grupo experimental.

2.2.3. Comprensión del nivel de significación y el p -valor

Un concepto central en inferencia estadística que suele tener una interpretación

incorrecta es el nivel de significación, denotado habitualmente como α o $\alpha\%$. Recordemos que el *nivel de significación* se define como la probabilidad de rechazar la hipótesis nula (H_0), supuesta cierta, es decir, $\alpha = P(\text{rechazar } H_0 | H_0 \text{ es cierta})$. Supongamos, por ejemplo, que trabajamos con un nivel de significación del 5%, esto significa que, si la hipótesis nula H_0 es cierta, se rechazará 5 de cada 100 veces, o lo que es lo mismo, en 100 veces que apliquemos un contraste de hipótesis utilizando este nivel de significación, rechazaremos una hipótesis nula cierta 5 veces (Birnbbaum, 1982; Vallecillos, 1999).

En relación a los errores asociados a la interpretación incorrecta del nivel de significación destacamos el que consiste en intercambiar los términos de la probabilidad condicionada que contempla su definición. Es decir, se interpreta como la probabilidad de que la hipótesis nula sea cierta sabiendo que se ha tomado la decisión de rechazarla, es decir, $\alpha = P(H_0 \text{ es cierta} | \text{se ha rechazado } H_0)$. Con respecto a dicho error encontramos, por ejemplo, las investigaciones de Birnbbaum (1982). El autor pidió a sus estudiantes que definiesen el nivel de significación, siendo la respuesta más frecuente: “Un nivel de significación del 5% significa que, en promedio, 5 de cada 100 veces que rechazamos la hipótesis nula estaremos equivocados”.

El mismo error también fue encontrado en la investigación de Falk (1986), donde la mayoría de sus estudiantes creían que el nivel de significación era la probabilidad de equivocarse al rechazar la hipótesis nula. Por su parte, Vallecillos (1994, 1999), señala que este tipo de error en más de 400 estudiantes de diferentes especialidades en la Universidad de Granada (por ejemplo, de medicina, ingeniería o psicología). En la investigación realizada, la autora analiza las respuestas dadas por los participantes a los dos ítems siguientes en su cuestionario:

Item 1: Un nivel de significación del 5% significa que, en promedio 5 de cada 100 veces que rechazamos la hipótesis nula estaremos equivocados (verdadero/falso). Justifica tu respuesta.

Item 2: Un nivel de significación del 5% significa que, en promedio, 5 de cada 100 veces que la hipótesis nula es cierta la rechazaremos (verdadero/falso). Justifica tu respuesta.

Sólo el 32% de los estudiantes de la investigación de Vallecillos (1994) dio una respuesta correcta al ítem 1 (opción falsa) y el 54% dio una respuesta correcta al ítem 2 (opción verdadera). De 135 estudiantes que justificaron su respuesta, el 41% dio un argumento correcto en los dos ítems. Resultados similares fueron encontrados por Krauss y Wassner (2002) en profesores de universidad responsables de la enseñanza de métodos de investigación y en Batanero, Vera y Díaz-Batanero (2012) con estudiantes de psicología.

Batanero (2000) resalta que la probabilidad de que la hipótesis nula sea cierta una vez la hemos rechazado y la probabilidad de que la hipótesis nula sea cierta una vez que hemos obtenido el valor del estadístico de contraste no pueden conocerse en la metodología frecuencial, sino sólo en la aproximación bayesiana. En consecuencia, los estudiantes que consideran correcto el ítem 1, planeado por Vallecillos, darían una interpretación bayesiana al resultado de una metodología frecuencial.

Interpretaciones similares se encuentran sobre el significado del p -valor. Algunos estudiantes piensan que el p -valor es la probabilidad de que el resultado obtenido al realizar un contraste estadístico se deba al azar (Batanero, 2000). Esta interpretación es también errónea, pues el p -valor es la probabilidad de obtener el resultado particular u otro más extremo cuando la hipótesis nula es cierta y no hay otros factores posibles que influyan el resultado (de la Fuente y Díaz-Batanero, 2004). Por consiguiente, no podemos inferir la existencia de una causa particular en un experimento a partir de un resultado significativo, a menos que la muestra sea aleatoria y el experimento tenga controladas todas las posibles variables que le puedan afectar. Estos errores los encontramos también en las investigaciones de Cañadas et al. (2012) con estudiantes de psicología donde, al resolver un contraste chi-cuadrado los participantes confunden, además, el nivel de significación con el concepto de p -valor.

Otro error común relacionado con los anteriores es la creencia en la conservación del valor del nivel de significación cuando se realizan contrastes consecutivos en el mismo conjunto de datos. Moses (1992) describió esta creencia como el *problema de las comparaciones múltiples*. Consiste en que al aplicar muchos contrastes de hipótesis al mismo conjunto de datos el nivel de significación real crece y ya no es el asumido por el investigador.

Usualmente los estudiantes que realizan una interpretación incorrecta del nivel de significación o del p -valor arrastran una interpretación incorrecta de los resultados significativos (Díaz-Batanero, 2018). Un resultado significativo solo establece la frecuencia relativa de veces que se rechazaría la hipótesis nula cierta a la larga (error tipo I). Pero se confunde significación estadística y práctica, es decir, se supone que un resultado estadísticamente significativo implica un efecto de gran tamaño. Sin embargo, con un pequeño tamaño del efecto y una muestra muy grande es fácil obtener resultados estadísticamente significativos (Batanero, 2000).

2.2.4. Comprensión del proceso de contraste

Otras dificultades señaladas en la literatura se refieren a la comprensión del razonamiento que subyace al contraste de hipótesis. Ello es debido al aparente parecido de este razonamiento con una demostración por reducción al absurdo donde se razona en la forma siguiente (de la Fuente y Díaz-Batanero, 2004; Vallecillos, 1999):

Si A, entonces B no puede ocurrir.

B ocurre; entonces deducimos que A es falso.

Por ejemplo, decimos todos los hombres tienen la sangre caliente. Si encontramos un ser vivo de sangre fría, no puede ser un hombre. Pero, en un contraste estadístico se razona en la forma siguiente:

Si A, entonces es muy poco probable el suceso B;

Ocurre B, y rechazamos A

Sin embargo, no sería una conclusión válida deducir que es muy improbable que A sea cierto. Por ejemplo, si una persona no está contagiada con la COVID-19, es poco probable que una prueba médica sea positiva. Sin embargo, si la prueba médica es positiva, esto no implica que la persona esté enferma, pues puede ser un falso positivo, que aparecen porque la misma prueba se aplica muchas veces. Es decir, un suceso muy improbable sucede a veces si se repite muchas veces el experimento (Vallecillos, 1999).

En relación a este tipo de razonamiento Vallecillos (1994) describió dos creencias erróneas: 1) identificación de un contraste de hipótesis como una prueba probabilística, creencia que es aplicada bajo el marco de la inferencia bayesiana; 2) considerar, una vez realizado el contraste de hipótesis, la veracidad o falsedad de una hipótesis. Falk y

Greenbaum (1995) consideran que este tipo creencia se debe a la existencia de mecanismos psicológicos.

2.3. COMPRESIÓN DEL INTERVALO DE CONFIANZA

A continuación, se resumen las investigaciones relacionadas con la comprensión del intervalo de confianza y sus propiedades (Behar, 2001; Cumming y Fidler, 2005; Cumming et al., 2004). Dicho procedimiento permite la estimación de los parámetros que caracterizan las distribuciones de las variables aleatorias y además de indicar qué valores se puede esperar para los parámetros y expresar la incertidumbre de la estimación (Cepeda-Cuervo et al., 2008). Su interés actual se debe a que asociaciones como la *American Psychological Association* recomiendan complementar los contrastes de hipótesis con el cálculo de los intervalos de confianza (Yaremko et al., 2013).

2.3.1. Definición del intervalo y del nivel de confianza

Behar (2001) evalúa la comprensión del intervalo de confianza de un grupo de 47 profesores, y otro de 297 estudiantes universitarios. Parte de los participantes en los dos grupos indicaron que los valores que constituyen un intervalo de confianza se refieren a la variable aleatoria o al estadístico que se usa como estimador y no al parámetro en estudio. También, en la investigación desarrollada por Olivo (2008), el autor señala que un 25% de los estudiantes confundía estadístico con parámetro y, además, consideraban que la construcción del intervalo se refiere al estadístico y no al parámetro. De forma similar, en la investigación llevada a cabo por Cumming y Fidler (2005) con 180 estudiantes de psicología, los autores mostraron que los intervalos de confianza eran considerados por los estudiantes como estadísticos descriptivos, ignorando su naturaleza inferencial. El 38% de los estudiantes consideraron que los intervalos de confianza proporcionan valores plausibles para la media de la muestra y el 19% confundían el intervalo de confianza con el rango.

Otro de los errores relacionados encontrados en la literatura, es en relación a la interpretación del nivel de confianza. El 40% de los expertos y 65% de los estudiantes en el trabajo de Behar (2001), lo interpretaron como la probabilidad de que el parámetro

se encuentre en el intervalo en lugar de asociarlo al porcentaje de intervalos calculados a partir de muestras de igual tamaño en la población que contiene al parámetro. Esta interpretación también fue hallada en el trabajo de Olivo (2008) con 252 estudiantes de ingeniería, quien sugiere, no se asocia el nivel de confianza con la frecuencia relativa a largo plazo con la que los intervalos generados aleatoriamente contengan al verdadero parámetro de la población. El autor indica que hay una interpretación bayesiana del intervalo de confianza, pues suponen que los extremos del intervalo son fijos y no aleatorios.

Por otro lado, en el estudio de Behar (2001) destaca que el 29% de los expertos y la mitad de los estudiantes asumieron el nivel de confianza como el porcentaje de datos poblacionales que están contenidos en el intervalo de confianza. Es decir, por ejemplo, se piensa que el intervalo del 95% contiene el 95% de los valores posibles de la población en estudio.

Por su parte Cumming et al. (2004) y Fidler y Cumming (2005) estudian sistemáticamente los errores en la interpretación de intervalos de confianza de los investigadores que han publicado artículos en los que usan dichos intervalos. En concreto analizaron la creencia en la replicación, es decir la predicción del nuevo intervalo de confianza si se replica el experimento. Para ello contactaron con 134 autores de artículos publicados en revistas internacionales que habían utilizado métodos estadísticos en sus trabajos. A los investigadores les proporcionaron un intervalo de confianza del 95% para la media de una población y les preguntaron cuáles serían los valores más probables de la media muestral si se replicase el mismo experimento un cierto número de veces. Los autores informan que la mayoría de los encuestados (78%) indicaron al construir otro intervalo de confianza, la probabilidad de que el parámetro poblacional caiga en el mismo intervalo es el 95% aunque los autores indican que esta probabilidad es de hecho del 83%.

Por su parte, Belia et al. (2005) enviaron, a través de correo electrónico, un cuestionario a un grupo de 473 investigadores de las disciplinas de psicología, neurociencia del comportamiento y medicina, solicitando que juzgaran cuando eran significativamente diferentes las medias de dos grupos independientes, a partir de la interpretación de los intervalos de confianza presentados gráficamente. Los autores

observaron que sólo unos cuantos de los participantes que completaron la actividad se acercaron a la respuesta correcta. Un grupo de los que contestaron interpretaron los resultados como si se refiriesen a una sola población, y otro resolvió el problema como si se tratase de dos medias relacionadas. Además, muchos investigadores creían que dos intervalos de confianza del 95% de confianza distintos para la misma media no se pueden solapar. También pensaban que cuando dos intervalos para medias independientes no se solapan, ello implica una diferencia significativa en el nivel de 0,05 entre las medias, y que el solapamiento de los dos intervalos implica que no existe una diferencia significativa.

Ninguna de estas creencias es totalmente cierta, pues depende de si las muestras son independientes o relacionadas. En el caso de muestras relacionadas, el intervalo de confianza de la diferencia es menor que la diferencia de los intervalos de cada media por separado luego podría haber algo de solapamiento entre los intervalos de cada media y ser la diferencia estadísticamente significativa. Los autores dan una regla para decidirlo que depende del tamaño de la muestra (al menos 10 elementos) y de la magnitud del solapamiento (que sea una proporción pequeña del intervalo).

La única investigación realizada con profesores es la de Yáñez y Behar (2009), quienes realizan un estudio con 15 futuros profesores de Educación Secundaria en Colombia, a los que realizan entrevistas. Una de las preguntas realizada fue si el intervalo de confianza contiene a la media muestral. Uno de los participantes no niega esta afirmación, aunque indica que el porcentaje de medias muestrales incluidas en el intervalo de confianza puede ser diferente al nivel de confianza y que ello dependerá de la diferencia entre la media muestral y la poblacional.

Roldán et al. (2020) realizan una investigación con estudiantes españoles de bachillerato que se preparaban para las pruebas de acceso a la universidad y habían estudiado previamente los intervalos de confianza. Entre otras dificultades encuentran que los estudiantes consideran los extremos del intervalo como valores constantes, dan una interpretación bayesiana al intervalo y piensan que el intervalo se refiere a la media muestral y no a la media poblacional. También suponen relación directa entre precisión y nivel de confianza y entre el ancho del intervalo y el tamaño de la muestra, aunque estas relaciones son en realidad inversas. Finalmente indican que algunos estudiantes

sitúan la media muestral en un extremo del intervalo y otros no comprenden la finalidad de la estimación por intervalos.

2.3.2. Propiedades del intervalo de confianza

En relación a las propiedades del intervalo de confianza, en varios estudios se pudo comprobar también una cierta debilidad conceptual respecto a la relación existente entre amplitud del intervalo, el tamaño de la muestra y el nivel de confianza (Behar, 2001; Fidler, 2005).

Behar (2001) señala la falta de comprensión, por parte de los participantes en del estudio, acerca de relación entre el ancho del intervalo y el nivel de confianza. Se cree que no es posible disminuir la confianza del intervalo, como si esta fuese inseparable del ancho del mismo. Otra parte de los estudiantes asumen incorrectamente que un nivel de confianza alto va unido exclusivamente a un intervalo más estrecho, en cualquier caso, sin tener en cuenta el tamaño de muestra o la varianza de la población. También se supone que el intervalo se refiere a los posibles valores de la variable aleatoria y no del parámetro. No se perciben los extremos del intervalo como aleatorios.

Garfield et al. (2004), desarrollan un módulo de enseñanza sobre el intervalo de confianza, indicando que se deben aprender las siguientes propiedades:

- Un intervalo de confianza es un conjunto de valores plausibles del parámetro que pudiera haber generado el dato observado.
- Se calcula mediante un estadístico muestral más o menos una medida del error del muestreo.
- El nivel de confianza nos da la probabilidad de que el método construya un intervalo que incluya el parámetro desconocido. La probabilidad se relaciona con el método, no con el parámetro. Si se recogen muchas muestras aleatorias independientemente de una población y se construyen los intervalos de confianza del 95% para cada muestra, en alrededor de un 5% de los intervalos no se incluye el parámetro poblacional. No sabemos si el parámetro está o no incluido en el intervalo.

- Un incremento en el tamaño de la muestra disminuye el ancho del intervalo, mientras que el nivel de confianza lo amplía.

También señalan las siguientes concepciones erróneas frecuentes entre los estudiantes:

- En un intervalo de confianza calculado con un nivel de confianza del 95%, hay un 95% de probabilidad de que el intervalo de confianza incluya la media muestral.
- En un intervalo de confianza calculado con un nivel de confianza del 95%, hay un 95% de probabilidad de que la media poblacional estará entre los dos valores (límite superior y límite inferior).
- Un intervalo ancho significa menos confianza.
- Un intervalo de confianza estrecho es siempre preferible.

Cumming y Fidler (2005) comprobaron las dificultades que existen sobre las relaciones que hay entre los distintos conceptos que intervienen en los intervalos de confianza indicando que no se comprende la relación de la amplitud del intervalo con diferentes factores; por ejemplo, se piensa que la amplitud del intervalo de confianza se incrementaría si se aumentara el tamaño de la muestra (Fidler y Cumming, 2005). Olivo (2008) describe la comprensión mostrada por 252 estudiantes de ingeniería. Entre los resultados obtenidos se destaca que el 33% no comprende la relación entre la amplitud del intervalo y el tamaño de muestra o nivel de confianza. Asimismo, apuntan que un 32,1% de estudiantes que han calculado correctamente el intervalo de confianza para una media, dan al mismo una interpretación bayesiana, como intervalo de credibilidad. Otro de los errores descritos en la construcción del intervalo, es usar la desviación típica de la población, en lugar de la desviación típica de la distribución muestral.

2.4. DIFICULTADES CON OTROS CONCEPTOS DE INFERENCIA

Para completar el análisis de las dificultades en la inferencia estadística, a continuación, analizamos las relacionadas con la distribución muestral y el teorema central del límite,

que son ampliamente utilizados para la construcción del intervalo de confianza y el contraste de hipótesis.

2.4.1. Comprensión de la distribución muestral

Al iniciar el estudio de la inferencia estadística, los conceptos de población y muestra, junto a sus características, adquieren una gran importancia. El proceso de muestreo, que permite obtener al azar un valor de la variable, debe comprenderse como un sistema de ideas interrelacionadas (selección al azar, replicación, variabilidad y distribución), para lo cual es esencial la identificación de la distribución en muestreo (Chance et al., 2004). Sin embargo, tal identificación genera ciertas confusiones, ya que es necesario manejar simultáneamente diferentes distribuciones (Batanero et al., 2019; Harradine et al., 2011).

El proceso de inferencia comienza considerando una población con una variable aleatoria y la *distribución de la variable en la población*; por ejemplo, la altura de chicos de 18 años en un cierto país o ciudad. En esta población interesa conocer un parámetro, por ejemplo, la altura media de todos los chicos en la población. Para estimarla se toma una muestra aleatoria de datos; por ejemplo 500 chicos a los que se mide su altura. A partir de estas medidas se obtiene la *distribución de la variable estadística*, “altura de los 500 chicos de la muestra” y la media de dicha muestra, \bar{x} , será empleada para estimar la media de la población. A partir del ejemplo empleado, es evidente que es necesario trabajar con dos medias diferentes; la de la población y la de la muestra. De acuerdo a la aproximación frecuencial de la inferencia estadística, la primera es una constante de la que no se conoce el verdadero valor de la media de la población, mientras que la segunda es un valor conocido en la muestra.

Ahora bien, cuando se considera no sólo la muestra utilizada, sino todas las posibles muestras de tamaño 500 tomadas aleatoriamente de la misma población, al calcular la media de cada una de estas muestras, se tiene una variable aleatoria, pues el valor de la media varía de una muestra a otra. La distribución de esta nueva variable aleatoria (media de todas las posibles muestras del tamaño dado) es la *distribución muestral*. En el caso particular del ejemplo tratado, la teoría estadística indica que la

distribución muestral es una distribución normal, cuya media coincidiría con la media de la variable de partida en la población.

En definitiva, tenemos tres distribuciones diferentes (distribución de la variable en la población, distribución de la variable estadística en la muestra y distribución muestral del estadístico) y tres medias diferentes, (parámetro, estadístico y estimador del parámetro o media de la distribución muestral) que habitualmente los alumnos confunden (Schuyten, 1991). El trabajo con estos tres niveles y el uso de cada uno en el proceso de inferencia tiene una gran complejidad semiótica, lo que explica que los alumnos cometan errores, como se muestra en algunas investigaciones.

Son muchas las investigaciones que dan cuenta de esta dificultad. Una de las primeras fue realizada por DelMas et al. (1999). Los autores utilizaron un software, para desarrollar con los estudiantes actividades de muestreo, partiendo de variables con diferentes distribuciones, como normal, uniforme o en forma de U. Los estudiantes podían cambiar los parámetros poblacionales y el tamaño de la muestra para simular la toma de muestras, calcular un estadístico en cada muestra y representar la distribución muestral del estadístico. Los autores resaltan la importancia de las siguientes propiedades, para comprender la distribución muestral:

- Relación entre tamaño muestral y la dispersión de la distribución muestral;
- Propiedades que caracterizan el estadístico como buen estimador del parámetro de la población;
- Comprobar el comportamiento de los estadísticos de la distribución muestral al crecer el tamaño de muestra, correspondiendo con lo previsto por el teorema central del límite.

En el trabajo desarrollado por Chance et al. (2004), los autores sugieren que el aprendizaje se refuerza cuando los estudiantes pueden comparar sus propias creencias con los resultados empíricos y que la simulación, en general, mejoró la comprensión de los estudiantes en su experiencia. A la misma vez, se destacó que el uso del ordenador supone un coste y esfuerzo añadido por parte de los estudiantes. Este hecho, entre otros, ha supuesto que se cuestione la efectividad del uso de métodos de simulación por ordenador (*Computer simulation methods*, CSMs). Sin embargo, existe una corriente de

investigadores que subrayan las ventajas que supone simular un número elevado de muestras de una población y construir, posteriormente, la distribución muestral empírica (Chance et al., 2004; Beckman et al., 2017) en el proceso de enseñanza de la distribución muestral. En esta línea, Lipson (1997), usando los programas Minitab y Sampling Laboratory trabajó las distribuciones muestrales con dos grupos de alumnos, pasando un cuestionario después de la enseñanza. La autora analizó los mapas conceptuales que hicieron los alumnos antes y después de las simulaciones con ordenadores y no encontró grandes diferencias entre los estudiantes que utilizaron distintos softwares, pero las actividades de simulación mostraron la mejora de la comprensión de la distribución muestral.

Liu y Thompson (2009) indican que los estudiantes sólo consideran la muestra dada en los ejercicios que resuelven, sin ser capaces de imaginar las infinitas muestras potenciales que pueden tomarse de una población; por ello no ven a la media de la muestra como una variable aleatoria. Esto es un obstáculo para comprender la idea de distribución muestral. Otros estudios relacionados con el tema son los llevados a cabo dentro del programa de heurísticas y sesgos en psicología (Kahneman et al., 1982). En este programa la *heurística* es una regla de conducta inconsciente, que guía la resolución de las tareas de razonamiento complejas, reduciendo la información de las mismas. Pero a veces fallan y producen sesgos en las decisiones bajo incertidumbre.

En relación con el muestreo, la *heurística de representatividad* consiste en considerar la probabilidad de obtener una muestra comparando la similitud que guarda la muestra con la población a la que pertenece, o bien con el proceso de muestreo que la ha generado. Por ello se espera que cualquier muestra, independientemente de su tamaño, sea un fiel reflejo de la población de donde ha sido extraída. Una de las creencias erróneas sobre este tema relacionado es la *insensibilidad al tamaño de la muestra*, donde no se tiene en cuenta dicho tamaño para estimar la variabilidad del muestreo. Los autores hablan de la “ley de los pequeños números”, cuando se espera una convergencia a la distribución teórica incluso en pocas repeticiones de un experimento, generalizando indebidamente la *ley de los grandes números* (Batanero et al., 2019).

Shaughnessy et al. (2004) analizan la comprensión de la distribución muestral en 272 estudiantes de entre 10 y 19 años, utilizando como contexto una máquina de donde se extraen al azar caramelos de dos colores (40% rojos y 60% negros), y pidiendo a los estudiantes estimar el número de caramelos rojos en una muestra de 10 y otra de 100 caramelos. Los autores informan que el 25% de los participantes esperaban obtener el mismo resultado en las dos muestras repetidas del mismo. Los autores identifican en los estudiantes tres tipos de razonamiento sobre la distribución muestral: a) aditivo, que consiste en considerar las diferentes muestras como subconjuntos disjuntos de la población; b) proporcional, en la que se comprende el valor esperado de la distribución muestral y c) distribucional, donde se comprende tanto el valor esperado como la variabilidad de la distribución muestral. La mayoría de los participantes muestran el razonamiento aditivo y proporcional y la variabilidad muestral se sobreestima, pues muchos de los estudiantes creen en la posibilidad de resultados muy variados, independientemente del tamaño de la muestra.

Begué et al. (2018) analizan la comprensión del valor esperado y la variabilidad de la distribución muestral de la proporción y el efecto del tamaño de la muestra sobre dicha variabilidad en 302 estudiantes de segundo y cuarto curso de Educación Secundaria Obligatoria. Piden para ello generar muestras de cuatro valores de distribuciones binomiales con diferentes parámetros. Concluyen que los estudiantes tienen una buena intuición del valor esperado, es decir, la proporción de resultados para el caso favorable en las muestras se acerca a la proporción teórica. No obstante, no es tan buena la comprensión de la variabilidad, sobreestimando la variabilidad de muestras grandes, por lo que las autoras concluyen que los alumnos no alcanzan el nivel de razonamiento distribucional sobre el muestreo. Estos resultados se confirmaron posteriormente al repetir la investigación con estudiantes de bachillerato (Batanero, Begué, Borovcnik y Gea, en prensa). Aunque estos estudiantes tienen mejor comprensión de la variabilidad muestral, todavía son muchos los que no alcanzan el razonamiento distribucional, lo que se confirma en el análisis de los argumentos proporcionados para justificar sus respuestas.

2.4.2. Comprensión del teorema central del límite

Otro punto importante en la aplicación de la inferencia estadística es la comprensión del teorema central del límite, puesto que, tanto para en el contraste de hipótesis como para el intervalo de confianza, se aplicará con objeto de determinar la distribución muestral de la media o proporción considerando muestras de tamaño suficientemente elevado que garantice la convergencia a la distribución normal. La comprensión de este teorema, según Castro-Sotos et al. (2007), implica aceptar que la distribución muestral se distribuye normalmente si el tamaño muestral tiende a infinito.

Basándose en su estudio de libros de texto, Alvarado (2007) diseña un experimento de enseñanza del teorema, aplicando el enfoque ontosemiótico y aplicándolo al campo de la ingeniería. Su experiencia de enseñanza incorpora el uso de la hoja Excel y el programa estadístico @risk para simular distribuciones de probabilidad en el estudio del muestreo. El autor compara los elementos del significado institucional pretendidos con los significados implementados y adquiridos por los estudiantes. Los resultados mostraron una buena comprensión en relación a:

- Aproximación de la distribución normal a la distribución binomial;
- No tuvo mayor dificultad el lenguaje simbólico, con algunas excepciones en los problemas algebraicos abiertos;
- El procedimiento algebraico y la tipificación, así como la simulación con apoyo de Excel;
- Al uso de las propiedades de la media y varianza de la suma de variables aleatorias.

Algunas dificultades que se observaron fueron: la identificación de la distribución de probabilidad que modela la variable estadística; expresión verbal incorrecta en una gama de conceptos relacionadas con el teorema; errores de tipificación; no aplicar la corrección de continuidad y no ser capaces de dar una decisión en base al análisis del teorema central del límite. Asimismo, los estudiantes erróneamente lo extrapolan a situaciones inadecuadas.

Méndez (1991) recomienda contar con material manipulable o bien simulación con ordenador para visualizar la extracción de muestras y la distribución muestral. El autor identifica cuatro propiedades esenciales en la comprensión del teorema:

- La media de la distribución muestral es igual a la media de la población, e igual a la media de una muestra cuando el tamaño de la muestra tiende al infinito;
- La varianza de la distribución muestral es menor que la de la población (cuando el tamaño muestral es superior a uno) y disminuye al aumentar el tamaño de la muestra;
- La forma de la distribución muestral tiende a ser acampanada a medida que se incrementa el tamaño muestral, y aproximadamente normal, independientemente de la forma de la distribución en la población;
- La forma de la distribución muestral incrementa en apuntamiento y decrece en dispersión a medida que el tamaño muestral crece.

2.5. ANÁLISIS DE PROBLEMAS DE INFERENCIA ESTADÍSTICA

Los libros de texto son un importante recurso en la enseñanza de cualquier materia, pues concretan el currículo pretendido, fijado en las directrices curriculares. Autores como Reynolds et al. (2004) resaltan que los libros de texto facilitan el aprendizaje de las matemáticas pues generalmente los contenidos son presentados en distintos contextos permitiendo al estudiante observar su utilidad. Cordero y Flores (2007) indican que el discurso matemático escolar es determinado con frecuencia por el libro de texto, que regula las acciones de enseñanza y aprendizaje, junto con las creencias de los profesores. Ocelli y Valeiras (2013) resaltan su papel como herramientas que median entre el currículo y el aula y que concretan los contenidos pretendidos en las directrices curriculares. El papel de los libros de texto en la determinación de los contenidos, conceptos y procedimientos que se imparten hace que sea importante analizar cómo se aborda el tema de la inferencia estadística en estas fuentes. Seguidamente, se revisará también los estudios que analizan los contenidos reflejados en las pruebas de acceso a la universidad.

2.5.1. Análisis de la inferencia estadística en libros de texto

Numerosos trabajos analizan la presentación de un tema particular en los libros de texto (Fan, 2013; Fan et al., 2013), aunque para el caso particular de la inferencia, el número de trabajos con este enfoque es menor. Rodríguez-Muñiz y Díaz (2018) analizan la investigación sobre estadística y probabilidad en libros de texto de bachillerato, constatando que, aunque esta investigación ha crecido mucho, “La inferencia estadística está infrarrepresentada [...] teniendo en cuenta su importancia en el currículo de Bachillerato LOE-LOMCE” (p.73).

Alvarado (2007) realiza un estudio de 16 textos dirigido a estudiantes de ingeniería describiendo, a partir de ello, nueve campos diferentes de problemas de donde surge el teorema central del límite. Indica que los libros de textos carecen de la mayoría de los campos históricos del teorema central del límite, teniendo prioridad los campos de problemas indirectos.

Olivo (2008) utiliza el enfoque ontosemiótico para analizar y clasificar los diversos objetos matemáticos asociados al intervalo de confianza en una muestra amplia de libros de texto universitarios. Describe una gran diversidad de campos de problemas, acompañados con una gran variedad de lenguaje, argumentos y propiedades. El autor señala conflictos semióticos en los textos en cuanto a la definición del intervalo; además en un muy alto porcentaje de libros se da prioridad a las definiciones y propiedades y no se tratan las aplicaciones.

García y García (2009) realizaron un estudio relacionado con los términos específicos de inferencia estadística en cuatro libros de texto de segundo curso de bachillerato de Ciencias Sociales cuando se comienza el estudio formal de dicho tema. Los autores realizan una clasificación de dichos términos en base al significado que presentan éstos en los contextos matemático y cotidiano. Se resalta el contexto y se presentan propuestas didácticas para superar los posibles problemas de aprendizaje.

En la misma línea, Ortiz et al. (2017) analizan el lenguaje matemático que se emplea para describir la estimación de la media e intervalos de confianza en los textos españoles de bachillerato. Concretamente los autores analizan las expresiones verbales, el lenguaje numérico, el lenguaje simbólico, el lenguaje tabular y gráfico. Entre los

resultados, los autores resaltan la variedad de expresiones verbales más específicas de estadística que de probabilidad. Aunque se destaca la gran riqueza y diversidad de expresiones verbales y predominio de lenguaje formal, se señala la importancia de la búsqueda de estrategias con objeto de afianzar un lenguaje matemático más avanzado y que permita realizar interpretaciones acertadas en relación a los significados más complejos que intervienen en la inferencia estadística.

2.5.2. Análisis de las pruebas de acceso a la universidad

Desde la implantación de las pruebas de acceso a la universidad son varios los trabajos que han abordado, desde diferentes puntos de vista, el estudio de los contenidos de las mismas. Ruiz et al. (2013), clasifican las distintas investigaciones que se han llevado a cabo sobre dichas pruebas en las siguientes líneas de investigación:

- Estudio de los resultados obtenidos por los estudiantes en dichas pruebas en algunas universidades, analizando las calificaciones en diferentes materias o comparando universidades entre sí;
- Comparación de las pruebas con otros métodos para evaluar a los alumnos en su entrada a la universidad en otros países;
- Estudios sobre los resultados y aspectos didácticos de alguna materia en particular de la prueba;
- Análisis de algunos de los factores que inciden en el rendimiento y la evaluación de los alumnos en las pruebas de acceso, así como los aspectos relativos a la validez y fiabilidad de las pruebas.

En este apartado se resumirán las principales investigaciones realizadas en esta línea. Si bien se enumeran trabajos en cada una de las líneas anteriores, pondremos especial énfasis en aquellas centradas en los resultados y aspectos didácticos. Puesto que la tesis se realiza en el área de didáctica de la matemática, se ha visto conveniente analizar los estudios llevados a cabo en la materia de matemáticas para posteriormente centrarnos en aquellos que tratan el tema de la inferencia estadística en las mencionadas pruebas.

2.5.2.1. Análisis de las pruebas de matemáticas

En la revisión sobre las investigaciones centradas en la asignatura de Matemáticas I y/o Matemáticas Aplicadas a las Ciencias Sociales II, se ha constatado distintos trabajos sobre esta temática. Con respecto a estas materias generalmente se han analizado los resultados de los estudiantes (Blázquez y Luengo, 1989; Cruz Sánchez, 2018), la opinión de los profesores (Murillo, 1997; Ruiz et al., 2013), las características psicométricas de las prueba (Gaviria, 2005), la evaluación de la calidad de la corrección de las pruebas (Cuxart, 2000; Grau et al., 2002; Mengual et al., 2013; Nortes et al., 2015) o las competencias matemáticas evaluadas (Boal et al., 2008).

Murillo (1997) entre los diversos estudios que realiza, muestra un análisis de las pruebas correspondientes a la convocatoria de junio durante el curso académico 1994-1995 de los 29 distritos universitarios. El autor señala que la asignatura Matemáticas II (actualmente denominada Matemáticas Aplicadas a las Ciencias Sociales II) es la asignatura que presenta mayor dificultad a los estudiantes desde el punto de vista del porcentaje de aprobados ya que más de la mitad de los estudiantes no llegaron a superar el examen. Por otro lado, en el trabajo se pone de manifiesto que la mayoría de las preguntas incluidas en las pruebas tienen un carácter práctico, y que los contenidos abordados son los asociados al cálculo de probabilidades y sistemas de ecuaciones lineales.

Contreras et al. (2010) analizan los significados de la integral definida en una muestra de libros de texto y en 140 pruebas de selectividad propuestas en la comunidad andaluza desde 1999 hasta 2008. A partir de dicho estudio, los autores obtienen una descripción de la realidad institucional con relación a la integral definida, que permite sugerir acciones para la mejora de los procesos de enseñanza y aprendizaje relacionado con el cálculo integral que se aborda en el segundo curso de Bachillerato de la asignatura Matemáticas I. Entre los resultados obtenidos, los autores destacan una reducción generalizada de la evaluación en las pruebas a métodos algebraicos y un aislamiento de las técnicas de resolución, indicando pues una baja idoneidad epistémica.

Nortes y Nortes (2010), analizan las respuestas de una muestra de alumnos a la materia Matemáticas Aplicadas a las Ciencias Sociales de las pruebas de acceso

realizadas en la Universidad de Murcia, en septiembre de 2009. En el trabajo, los autores analizan las calificaciones, detectan los errores más significativos y comparan los estadísticos de cada una de las cuestiones y de los bloques.

Ruiz et al. (2013) realizan un análisis estadístico de los resultados individuales obtenidos en Matemáticas por los alumnos en el País Vasco en el curso 2009-2010, comparando estos resultados por centro y analizando la relación de estos resultados con la opinión de los profesores sobre los mismos y la metodología del proceso de enseñanza-aprendizaje utilizada. Encontraron unos resultados muy pobres con pocas diferencias entre centros educativos. Un resultado no esperado es que los centros con buenos resultados utilizaban una metodología tradicional, con poca presencia de las nuevas tecnologías.

Mallart (2014) analiza las respuestas de 104 estudiantes en la asignatura de Matemáticas de las pruebas de acceso a la universidad de 2012, estudiando las resoluciones con el propósito de detectar errores significativos y poder incidir en tales dificultades de aprendizaje. Diferencia entre resoluciones mecánicas, la ejecución con precisión de las operaciones, la aplicación con rigor de las propiedades y la capacidad creativa de resolución. El estudio realizado muestra que las primeras son más frecuentes, lo que indica que los alumnos han aprendido los aspectos rutinarios en la resolución de problemas, pero se observa falta de creatividad, de rigor y precisión en los problemas geométricos.

Cabe señalar que pocos trabajos han estado centrados en el análisis de los problemas planteados. En esta línea, Ruíz et al. (2011) analizan los ejercicios de las PAU desde 1994 hasta 2008 en la Universidad del País Vasco con objeto de establecer una clasificación sobre los problemas tipo análogos en cuanto a su método de resolución de las materias de Matemáticas II y Matemáticas Aplicadas a las Ciencias Sociales II. Concretamente, los autores realizan una clasificación según el nivel de dificultad (estándar y difícil) y según el contenido tratado en el mismo. Los autores comprueban que un alto porcentaje de problemas de tipología difícil estaban relacionados con una calificación baja y un número alto de problemas de tipología estándar estaban asociados a calificaciones buenas. De los resultados, los autores recomiendan que, si se incluyeran

más problemas estándar en la materia de Matemáticas de Ciencias Sociales II, seguramente mejorarían los malos resultados que tienen los estudiantes en esta prueba.

En la misma línea, Díaz et al. (2014) realizan un estudio comparativo de cómo las pruebas PAU de cuatro Comunidades Autónomas tratan los contenidos de estadística y probabilidad, basándose en los contenidos curriculares de estos temas que se usan como criterio en su elaboración. Examinan en primer lugar estos contenidos, concluyendo que la distribución de las unidades curriculares en los exámenes es muy heterogénea. También examinan el contexto, indicando que en algunas comunidades aparecen ejercicios descontextualizados. En general, las pruebas priman un ejercicio de cálculo de probabilidades y otro de inferencia estadística.

Cruz Sánchez (2018) realiza un estudio sobre los resultados obtenidos en la asignatura de Matemáticas II en las pruebas de acceso a la universidad en Andalucía durante el periodo 2007-2017 (a excepción de 2010). Del estudio detecta que los problemas que no obtienen una buena calificación corresponden al bloque de análisis y el bloque de geometría. A partir de los resultados obtenidos y tras un análisis de las principales dificultades y errores que tienen los estudiantes cuando trabajan la geometría en la etapa de bachillerato, propone actividades de actuación con el fin de mejorar el proceso de enseñanza-aprendizaje del tema en cuestión.

Por parte del Grupo de Investigación sobre Educación Estadística de la Universidad de Granada, se han realizado diversos trabajos enfocados en el análisis de los problemas de probabilidad propuestos en las pruebas de acceso a la universidad en Andalucía en el periodo 2003-2014 (ej., Carretero et al., 2015; Contreras et al., 2015; López-Martín et al., 2015; López-Martín et al., 2016). Los resultados obtenidos en estos trabajos mostraron un fuerte énfasis en la probabilidad condicional, frente a la simple o compuesta, y en general, revelaban la dificultad de los problemas, con predominio de espacios muestrales compuestos, sucesos no equiprobables y datos dados en forma de probabilidad o porcentaje.

Parte de la presente memoria tiene como finalidad complementar los trabajos enfocados al análisis de los problemas de las pruebas de acceso a la universidad, para ofrecer una visión de la evaluación dedicada al bloque de estadística y probabilidad en las mencionadas pruebas en la Comunidad de Andalucía.

2.5.2.2. Estudio de la inferencia estadística en las pruebas de acceso a la universidad

Tal y como se ha puesto de manifiesto anteriormente, el número de investigaciones relacionadas con el presente trabajo es limitado. Las investigaciones más cercanas al análisis de los problemas de inferencia estadística en las pruebas de acceso a la universidad son las llevadas a cabo por Espinel et al. (2006), Espinel et al. (2007) y García y García (2005).

En el trabajo presentado por García y García (2005), los autores analizan las respuestas dadas por 50 estudiantes a los problemas de inferencia estadística de la convocatoria de junio de 2002 en la Universidad de la Laguna. Concretamente, los autores cruzan la solución del profesorado corrector con la forma de llevar a cabo el planteamiento, los cálculos y la resolución los estudiantes. Los resultados muestran que aproximadamente un 72% de los estudiantes seleccionan los dos ejercicios de inferencia estadística, sin embargo, solo siete alumnos consiguen concluir correctamente el contraste de hipótesis y determinar el intervalo de confianza.

Por otra parte, Espinel et al. (2006) seleccionan 272 estudiantes que realizaron la PAU en la Universidad de la Laguna en 2005 con el objeto de analizar la tendencia que tienen los estudiantes de elegir las actividades relacionadas con estadística y probabilidad. De los resultados se observa que las preguntas con contenidos estadísticos son las más seleccionadas por los estudiantes (90%) y corresponden a las calificaciones más elevadas, en contraposición con los ejercicios de funciones pues son seleccionados por un 20% de estudiantes y tiene asociados a las calificaciones más bajas.

Empleando la misma muestra, Espinel et al. (2007) ampliaron el estudio realizando una revisión sobre los errores más comunes cometidos por los estudiantes al abordar cuestiones de inferencia estadística. Los autores señalan que los estudiantes no han podido alcanzar la calificación máxima por no llegar a una solución correcta debido a los errores: dificultad al entender la noción de probabilidad, tipificación errónea de la distribución normal, uso incorrecto de la tabla de probabilidades de la normal, fallo en la aproximación de la binomial a la normal, confusión entre muestra y población, descuido ante cambios del tamaño muestral, dificultad para discernir entre parámetro, estimador y estimación, identificación y formulación confusa de los

contrastes unilaterales y bilaterales, intercambio del valor empírico con el valor muestral, asignación incorrecta del nivel de confianza, asociación errónea del nivel de significación y la probabilidad buscada, interpretación errónea de las conclusiones en los contrastes. Comparando los resultados con investigaciones previas, los autores señalan que el método de enseñanza-aprendizaje de la estadística no se ha adaptado lo suficientemente siendo necesario incluir ciertas mejoras. Por tal motivo, los autores proponen un desarrollo metodológico basado en: 1) uso de analogías en el desarrollo de los contenidos teóricos; 2) manejo de las TICs; 3) trabajo con proyectos estadísticos.

En resumen, observamos que, en las investigaciones halladas, los autores ponen el foco de atención en la forma de actuar de los estudiantes y en el estudio de las dificultades asociadas a la comprensión de conceptos estadísticos, dejando de lado la tipología de los problemas de inferencia y las variables que los definen. En base a ello, la realización del Capítulo 3 de la presente memoria de tesis tiene, entre uno de sus objetivos, aportar información sobre estas dos últimas cuestiones.

2.5.3. Análisis de otras pruebas de evaluación

Entre los investigadores que se han interesado por analizar otras pruebas de evaluación encontramos a Rico (2006a), quien relaciona el marco teórico de las pruebas de matemáticas PISA del año 2003 con la resolución de problemas desde el punto de vista curricular, a través del concepto de competencia matemática, utilizando un modelo funcional sobre el aprendizaje de las matemáticas. Dicho modelo analiza el patrón a seguir a la hora de resolver una tarea matemática: campos de actuación (fenómenos, situaciones o contextos y problema que plantea la tarea); herramientas o matemáticas puestas en juego para resolverla (estructuras conceptuales y procedimientos) y procesos cognitivos que se movilizan para su ejecución (pensar y razonar, comunicar, justificar, representar, modelizar, plantear y resolver problemas).

Caraballo (2010) analiza 173 ítems incluidos en las pruebas de diagnóstico correspondientes al segundo curso de Educación Secundaria Obligatoria de cinco comunidades autónomas durante el curso académico 2008-2009. La investigación se enfoca en estudiar la correspondencia entre los instrumentos de evaluación utilizados por las comunidades autónomas para realizar estas pruebas y el modelo PISA. En cada

uno de los problemas se analizó las variables: contenido matemático; contexto o situación; nivel de complejidad y procesos necesarios para realizar una tarea matemática. Caraballo (2010) destaca el desequilibrio existente en la distribución de los ítems respecto al contexto y al contenido. Además, observa que habitualmente la información viene dada mediante representaciones gráficas o simbólicas.

Por otro lado, Castellanos (2013) realiza un estudio sobre las tablas y gráficos estadísticos de las pruebas SABER de Colombia enfocados a estudiantes de último grado del ciclo de educación básica primaria en los años 2003, 2006 y 2009. La autora analiza el tipo de gráfico empleado o tabla empleada (como por ejemplo el diagrama de sectores o el pictograma); la competencia evaluada (razonamiento, resolución o comunicación); el nivel de lectura de gráficos (tipo de dato que debe obtener del gráfico); la actividad solicitada (interpretar una representación, organizar y representar datos) y los niveles de complejidad semiótica del gráfico (entre los descritos por Arteaga, 2011). En la misma línea, pero centrándose en las Pruebas de Diagnóstico Andaluzas Obligatorias para niños de 10 años, Mingorance (2014) analiza los gráficos y tablas estadísticas observando una alta frecuencia de los gráficos de barras frente a la escasez de otros gráficos recomendados en el currículo.

2.6. CONCLUSIONES

En todo proceso de enseñanza-aprendizaje, el conocimiento de las dificultades y errores cometidos por los estudiantes es un elemento que adquiere gran importancia. Esta información es de gran utilidad para el docente ya que además de saber cómo asimilan y abordan los alumnos un problema, le permitirá llevar a cabo actuaciones de mejora en la enseñanza. En el caso de la inferencia estadística, esta mejora traerá consigo el perfeccionamiento en sus aplicaciones futuras.

El estudio de los antecedentes nos revela una gran cantidad de investigaciones sobre los errores en la comprensión y aplicación de la inferencia estadística, que no sólo afectan a los estudiantes, sino incluso a profesores de estadística e investigadores. Además, nos informa de un extenso debate sobre la necesidad de mejora de la práctica estadística, así como de la enseñanza de esta materia.

Pocos de estos trabajos han sido realizados con futuros profesores de Educación Secundaria Obligatoria y Bachillerato, algunos de los cuáles podrían compartir las dificultades identificadas en otros colectivos. Por otro lado, además del conocimiento común de la materia (que sería el conocimiento que el profesor deberá enseñar a sus futuros estudiantes), los modelos de conocimiento del profesor de Matemática incluyen otros componentes. Entre ellos, en nuestro trabajo nos interesaremos por la faceta cognitiva de su conocimiento didáctico, en particular, su conocimiento de las posibles dificultades de los estudiantes. Todo ello justifica la evaluación que llevaremos a cabo en el segundo de nuestros estudios empíricos (Capítulo 4).

Para sintetizar nuestras conclusiones, a continuación, se identifican los conflictos semióticos subyacentes en las investigaciones analizadas, y algunas de las posibles causas que los provocan. Para finalizar, se muestran las conclusiones relacionadas con las investigaciones que analizan las pruebas de acceso a la universidad.

2.6.1. Conflictos identificados en la literatura

Respecto al nivel de significación (y el p -valor) se pueden encontrar numerosos conflictos semióticos. El más frecuente, tal y como se ha puesto de manifiesto en la Sección 2.2, está relacionado con el hecho de intercambiar, en su definición, los términos de la probabilidad, es decir, se considera que el nivel de significación es la probabilidad de que la hipótesis nula sea cierta sabiendo que esta ha sido rechazada. Este error se ha encontrado tanto en investigaciones en los que los participantes eran estudiantes (Birnbbaum, 1982; Falk, 1986; Vallecillos, 1994) como en aquellas en las que se contaba con la participación de profesorado del ámbito universitario encargados de la enseñanza de métodos de investigación (Krauss y Wassner, 2002). En el presente trabajo, este tipo de error es interpretado como conflicto semiótico cognitivo y subyace el error consistente en asignar a la probabilidad condicional la propiedad conmutativa (se pueden intercambiar sus términos).

Un segundo conflicto es la confusión entre la hipótesis nula y la hipótesis alternativa. Es decir, al establecer las hipótesis de un contraste, los estudiantes piensan que el fin de dicho procedimiento es demostrar la hipótesis nula en lugar de considerar

que es esta la que se pretende rechazar. Por otro lado, se confunden hipótesis unilaterales y bilaterales, se plantean hipótesis que no cubren el espacio paramétrico y se utiliza el estadístico muestral en vez del parámetro al plantear las hipótesis (Batanero et al., 2012; Cañadas et al., 2012; Vallecillos, 1999).

Vallecillos (1999) describe tres tipos de creencia errónea en relación a la naturaleza de la prueba proporcionada por el contraste de hipótesis:

- *El contraste de hipótesis en la metodología frecuencial es una prueba probabilística de la hipótesis*, que permite calcular la probabilidad de que una hipótesis sea cierta.
- *El contraste estadístico es un método matemático; como tal, y al ser la matemática una ciencia exacta, al finalizar hemos probado la verdad o falsedad de una hipótesis*. Esta creencia, siempre errónea, supone un conflicto entre una demostración deductiva y el contraste de hipótesis.

Respecto a los intervalos de confianza, el principal conflicto es interpretar sus extremos como fijos y no variables. Asimismo, se hallan conflictos relacionados con la definición del coeficiente de confianza y con la suposición de que el intervalo se refiere a valores de la variable aleatoria y del estadístico muestral y aquellos en los que no se relaciona el ancho del intervalo con diferentes factores.

Finalmente, en relación con la distribución muestral, se ha detectado confusión respecto a la distribución de la población o incluso con la distribución de datos en la muestra. Por otro lado, pocos estudiantes alcanzan el nivel de razonamiento distribucional en su comprensión de la distribución muestral.

2.6.2. Posibles causas de los conflictos descritos

Algunos de los conflictos presentados en la sección 2.6.1. son transmitidos de forma no intencionada durante el proceso de enseñanza de la inferencia estadística. Generalmente, el proceso de enseñanza-aprendizaje es demasiado formal, y se reduce en la mayoría de los casos al aprendizaje de las fórmulas de cálculo, sin dedicar suficiente tiempo a la interpretación de los resultados.

En este sentido, algunas propuestas actualmente tratan de desarrollar lo que se denomina *razonamiento inferencial informal* con poco rigor. Consiste en sustituir el empleo de la distribución muestral basada en el teorema central del límite por la simulación con ordenador (*Computer simulation methods*, CSMs). Además, no se introducen en forma teórica conceptos como error y sus tipos, región crítica y de aceptación y otros. Borovnick (2019) analiza lo que él denomina “enfoque informal de la inferencia” que consistiría en recurrir a la simulación para facilitar la comprensión, pero sólo como un paso intermedio al estudio más completo de la inferencia. Por otro lado, analiza la “inferencia informal” que considera no es verdadera inferencia, en cuanto se transmiten en la misma algunos posibles conflictos.

En el trabajo de Batanero et al. (2017, p. 12-13) las autoras describen una lista de algunos ejemplos de los conflictos en la enseñanza:

- *Confusión entre frecuencia relativa y probabilidad.* Es frecuente encontrar en la literatura las expresiones probabilidad empírica o probabilidad frecuencial para referirse a la estimación frecuencial de la probabilidad a partir de la frecuencia relativa. Este conflicto ya fue denunciado por Chaput, Girard y Henry (2011).
- De la misma forma se encuentra en la literatura expresiones como *probabilidad teórica* para contraponerla con la frecuencial o subjetiva, mientras que toda probabilidad es teórica; lo empírico sería su estimación por medio de la frecuencia. Es igualmente desafortunado el término *probabilidad experimental*, puesto que todas las probabilidades son teóricas.
- *Confusión entre la distribución muestral teórica y empírica.* Es importante que el profesor comprenda los límites y no sólo las posibilidades de la simulación. La simulación introduce un error de estimación añadido a los errores tradicionales en el estudio del contraste de hipótesis (errores tipo I y II). Además, sólo podemos controlar este error, a medida que aumenta considerablemente el número de simulaciones, pues la distribución muestral empírica, obtenida mediante simulación, converge a la teórica cuando el número de muestras tomadas tiende a infinito. Sin embargo, para un número finito de simulaciones, las distribuciones muestrales empíricas varían de una simulación a otra. Luego, si cada alumno simula su propia distribución muestral, los resultados obtenidos por cada uno de ellos pueden ser diferentes, lo que puede añadir complejidad a la clase. Una forma de solventar este problema sería que el profesor acumule los resultados de todas las simulaciones para que la clase trabaje con una única distribución muestral más precisa.
- *Confusión entre probabilidad e inferencia estadística.* En algunos trabajos sobre inferencia informal se plantean preguntas de tipo probabilístico, se consideran tales preguntas como tareas propias de la inferencia estadística, aunque falten los elementos mínimos que las caracterizan.
- En el caso del contraste de hipótesis, se simplifica excesivamente, sin considerar los posibles errores asociados, y se puede transmitir una visión muy determinista del contraste de hipótesis.
- En ocasiones no se diferencia entre la distribución estadística de datos en la muestra y la distribución de probabilidad en la población (detrás de las cuales

hay dos objetos matemáticos diferenciados, variable estadística y variable aleatoria), sino se considera que se trata de dos puntos de vista diferente sobre la misma distribución.

- Otras veces no se resalta la naturaleza condicional del p -valor, es decir, no se relaciona la probabilidad adecuadamente con la hipótesis supuesta y se produce de nuevo el conflicto consistente en intercambiar los términos en la probabilidad condicional al trabajar con el p valor y el nivel de significación. Al no explicitar los conceptos subyacentes, estos quedan sin fijarse en la mente de los alumnos.
- En otras propuestas de “inferencia informal”, simplemente se pide tomar una decisión subjetiva sobre el rechazo de una hipótesis utilizando sólo procedimientos descriptivos; lo cual no sería una verdadera inferencia estadística. En esta situación encontramos un conflicto entre la inferencia y análisis exploratorio de datos.
- Finalmente, en algunos casos, se trata de sustituir todo el razonamiento probabilístico mediante la tecnología, sin enfatizar en los procesos aleatorios involucrados en la simulación (Borovcnik, 2019). Tampoco esto sería una inferencia estadística, pues esta debe estar justificada por un modelo de probabilidad que ligue los datos a la población (Rossman, 2008).

Por otro lado, las autoras indican que en vez de enseñar inferencia estadística se enseña la construcción de modelos de simulación en un software específico. Esto añade tiempo a la enseñanza de dicho contenido, pues, salvo problemas muy triviales, la construcción de estos modelos entraña cierta complejidad. Otro conflicto, es la mezcla de las metodologías descrita por Fisher, Neyman-Pearson y bayesiana (Batanero et al., 2017; Gigerenzer, 1993).

2.6.3. Conclusiones sobre las investigaciones relacionadas con las pruebas de acceso a la universidad

Las pruebas de acceso a la universidad son un elemento fundamental para aquellos estudiantes que desean continuar sus estudios en la universidad. Además, en gran medida, son las que dirigen los contenidos a abordar durante la enseñanza en Bachillerato.

Aunque a lo largo de los años dichas pruebas han sufrido diversos cambios legislativos siempre ha tenido como objetivo confirmar si el futuro estudiante universitario ha adquirido las aptitudes, madurez académica y posee los conocimientos necesarios para acceder a las enseñanzas universitarias. El estudio de Rodrigo et al. (2012) refleja que la calificación obtenida por el estudiante junto a la modalidad de bachillerato y el orden de preferencia en la prescripción universitaria está relacionada

con el abandono de los estudios universitarios. Puesto que la calificación permitirá al estudiante seleccionar los estudios acordes con sus preferencias, los docentes procuran asegurar que los contenidos abordados en el curso de segundo de Bachillerato sean los evaluados en dichas pruebas verificando una correcta correspondencia entre las directrices curriculares y los contenidos de los que serán evaluados los futuros estudiantes universitarios.

La resolución de un problema de inferencia plantea un reto importante a los estudiantes y determinar claramente cuál es el contenido de los problemas sobre el tema en las pruebas de acceso a la universidad también lo es. Este tipo de análisis adquiere cierta importancia ya que el conocimiento de los contenidos puede promover la obtención de una mejor calificación en dicha prueba y, como consecuencia, facilitar el acceso del estudiante a los estudios universitarios. Además, el estudio y análisis de las pruebas de acceso a la universidad será de gran ayuda para el profesor de Bachillerato y al personal encargado de la elaboración de las pruebas.

La información recogida en este capítulo será de gran utilidad con el fin de poder analizar los contenidos de los problemas de inferencia estadística incluidos en las pruebas de acceso y conseguir así, establecer una idea inicial sobre la dificultad que estas tienen para los estudiantes. De la revisión bibliográfica sobre el análisis de las pruebas de evaluación se deduce la importancia que tiene la presentación de este trabajo con el fin de analizar si las pruebas están correctamente diseñadas para el fin con el que han sido propuestas.

Capítulo 3

LA INFERENCIA ESTADÍSTICA EN LAS PRUEBAS DE ACCESO A LA UNIVERSIDAD

Índice de Contenidos

- 3.1. Introducción
 - 3.2. Objetivos del Estudio 1
 - 3.3. Hipótesis o expectativas iniciales
 - 3.4. Metodología
 - 3.4.1. Descripción general de la metodología
 - 3.4.2. Muestra de problemas analizados
 - 3.4.3. Variables analizadas
 - 3.5. Campos de problemas, prácticas matemáticas y configuración de objetos matemáticos asociados
 - 3.5.1. Análisis semiótico de problemas propuestos
 - 3.5.2. Síntesis del análisis de los campos de problemas
 - 3.5.3. Estudio comparativo entre el significado pretendido y evaluado
 - 3.6. Modelo probabilístico de la población de partida
 - 3.7. Parámetro poblacional objeto de estudio
 - 3.8. Contexto del problema
 - 3.9. Análisis estadístico de los problemas
 - 3.9.1. Campos de problemas
 - 3.9.2. Modelo probabilístico de la población de partida
 - 3.9.3. Parámetro poblacional objeto de estudio
 - 3.9.4. Contexto del problema
 - 3.9.4.1. Estudio comparativo entre el campo de problema y el contexto
 - 3.10. Conclusiones del Estudio 1
 - 3.10.1. Idoneidad didáctica de los problemas de inferencia estadística en las pruebas de acceso a la universidad
 - 3.10.1. Conclusiones sobre los objetivos
 - 3.10.2. Conclusiones sobre las hipótesis
-

3.1. INTRODUCCIÓN

En este capítulo se describe el primero de los estudios empíricos llevados a cabo como parte de la tesis doctoral, que consiste en un análisis de los problemas de inferencia estadística propuestos en las pruebas de acceso a la universidad (PAU) en Andalucía en el período 2003 a 2014.

La enseñanza matemática está fuertemente condicionada por las pruebas de evaluación externas, pues es el deseo, tanto de los mismos estudiantes, como de los padres y profesores, que cada centro obtenga buenos resultados en ellas. De entre las pruebas de evaluación externas más relevantes por su impacto en la comunidad educativa, destacamos, el proyecto PISA (*Program for international Student Assessment*; OCDE, 2015), la evaluación TIMSS (*Third International Mathematics and Science Study*; Mullis et al., 2012), así como las pruebas de diagnóstico y las pruebas de acceso a la universidad (PAU).

Los resultados ofrecidos, tanto por las pruebas PISA (MCD, 2012; ME, 2009), como por las pruebas TIMSS (Mullis et al., 2012), sugieren que, para muchos estudiantes, las matemáticas solamente son un conjunto de reglas y técnicas sin significado y sin relación con la vida cotidiana. Los bajos resultados de los estudiantes españoles que evidencian estas pruebas hace reflexionar e investigar sobre los factores que pueden ser causa de ese fracaso.

En el caso de los estudiantes de Bachillerato, los contenidos recogidos en las orientaciones curriculares vienen indirectamente determinados por las PAU. El artículo 7 del Real Decreto 1892/2008, de 14 de noviembre, por el que se regulan las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades públicas españolas, establece que:

1. La prueba se adecuará al currículo del Bachillerato y versará sobre las materias, a las que se refieren los artículos 6 y 7 del Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del Bachillerato y se fijan sus enseñanzas mínimas, establecidas para el segundo curso.
2. Las Administraciones educativas y las universidades organizarán la prueba y deberán garantizar su adecuación al currículo del Bachillerato, así como la coordinación entre las universidades y los centros que imparten Bachillerato para la organización de la misma. (MP, 2008, p. 46934).

Aunque inicialmente la prueba está enfocada en la evaluación de los conocimientos y capacidades adquiridas por los futuros universitarios durante su etapa de Bachillerato, en la práctica se utiliza para seleccionar los estudiantes que desean ingresar en determinados estudios y centros universitarios.

Puesto que el objetivo principal de las PAU es evaluar el rendimiento que han adquirido los estudiantes cuando finalizan sus estudios de Bachillerato, el análisis y estudio de las mismas nos aportará una información valiosa sobre la adecuación de los contenidos de estas pruebas. Para estudiar si los contenidos de inferencia estadística de las pruebas están directamente relacionados con los recogidos en el currículum de Bachillerato de Humanidades y Ciencias Sociales, y el peso relativo que reciben dichos contenidos en el total de la prueba, se analizan los problemas de inferencia estadística formulados en las PAU cuando ha estado vigente el anterior currículum (MEC, 2007). Tal y como se mencionó en el Capítulo 1, la implementación de la LOMCE supuso la eliminación del contraste de hipótesis recogidos en la LOE, manteniéndose el resto de contenidos. Dado que el currículum en el marco de la LOMCE (MECD, 2015b) incluye criterios de evaluación y estándares de aprendizaje evaluables ligados específicamente al tema, que serán referentes en la planificación de la concreción curricular y en la programación didáctica, los resultados que se obtienen de este trabajo pueden servir también para planificar la enseñanza según esta normativa y preparar a los estudiantes para las pruebas finales de Bachillerato recogidas en la LOMCE.

La estructura del capítulo es la siguiente: en primer lugar, se presentan los objetivos e hipótesis del Estudio 1, seguidos de la descripción de la metodología que ha sido utilizada como base del estudio. Esta descripción incluye las características de la muestra empleada y de cada una de las variables que han sido objeto de estudio en este trabajo. Seguidamente, se describen cada una de las variables y categorías analizadas, presentando ejemplos de las mismas que faciliten su comprensión. Siguen los resultados del análisis estadístico de estas variables y las conclusiones más destacadas sobre la idoneidad didáctica de los problemas de inferencia propuestos en las PAU. Algunos resultados de este estudio se han publicado en Batanero, López-Martín, Arteaga y Gea (2018), López-Martín, Batanero, Díaz-Batanero y Gea (2016), López-Martín, Batanero, Gea y Arteaga (2016) y presentado en diferentes congresos que se recogen en el Apéndice.

3.2. OBJETIVOS DEL ESTUDIO 1

Como se ha indicado, el objetivo de este estudio fue analizar el contenido de los problemas de inferencia estadística propuestos en las PAU del Distrito Andaluz. Concretamente, se han seleccionado las pruebas de la especialidad de Bachillerato de Humanidades y Ciencias Sociales propuestas en el período desde 2003 a 2014 en la asignatura de Matemáticas Aplicadas a las Ciencias Sociales II. Se trata de identificar las principales variables que caracterizan estos problemas y analizar su distribución a lo largo del periodo seleccionado. Con este análisis se realiza una caracterización de los problemas relacionados con inferencia estadística incluidos en las citadas pruebas, con el fin de determinar el significado institucional de la inferencia estadística evaluado en las mismas y, poder así, comparar con el significado institucional pretendido en las orientaciones curriculares (Godino y Batanero, 1994). Este objetivo corresponde al objetivo específico OE2 de la presente tesis doctoral, enunciado en el Capítulo 1:

OE2. Caracterizar los problemas de inferencia estadística propuestos en las pruebas de acceso a la universidad y comparar el significado evaluado en dichas pruebas con el significado institucional pretendido para la inferencia estadística en las orientaciones curriculares.

Para alcanzarlo, se han identificado los problemas propuestos en las PAU y, mediante un análisis de su contenido se han identificado los campos de problemas propuestos, e identificado las configuraciones epistémicas que aparecen reflejadas en cada uno de los campos de problemas. Posteriormente, se han definido una serie de variables que han sido consideradas relevantes para la formación del estudiante en relación a los problemas de inferencia estadística y que se han deducido del estudio de la investigación previa. En concreto, se analizan los modelos teóricos de la población de partida y parámetro estimado. Por último, con objeto de valorar la idoneidad emocional y así analizar la conexión entre los conceptos aprendidos y las situaciones de la vida real, se ha examinado el contexto considerando los recogidos en las pruebas PISA (OCDE, 2015).

A partir de estas consideraciones, el objetivo del trabajo puede desglosarse en tres objetivos específicos:

OE2.1 Analizar los campos de problemas sobre inferencia estadística propuestos en las PAU, identificando los objetos matemáticos requeridos en su solución.

El interés de este objetivo radica en que los problemas propuestos a los estudiantes en las pruebas de evaluación van a incidir directamente en el aprendizaje de diferentes objetos matemáticos y el significado asignado a los mismos durante su formación. De acuerdo al enfoque ontosemiótico (Godino, 2002; 2017; Godino y Batanero, 1994; 1998; Godino et al., 2007), los objetos matemáticos surgen de las prácticas matemáticas realizadas para resolver problemas relacionados con el objeto.

Para lograr este objetivo y en base al enfoque ontosemiótico, se han resuelto cada uno de los problemas de las PAU para Matemáticas Aplicadas a las Ciencias Sociales II durante el periodo 2003-2014 con el fin de analizar los objetos relacionados con la inferencia estadística y determinar el significado evaluado en las pruebas. Se presenta en el capítulo un ejemplo de la resolución de un problema de cada uno de los campos de problemas identificados, identificando la configuración de objetos matemáticos ligada a su resolución.

OE2.2. Identificar las variables que definen los problemas de inferencia estadística propuestos en las PAU y analizarlas estadísticamente.

La investigación sobre comprensión de la inferencia estadística resumida en el Capítulo 2 muestra numerosas dificultades e indirectamente algunas variables de los problemas de inferencia estadística que inciden directamente en su dificultad. Es por ello importante identificar las principales variables que aparecen en los problemas propuestos a los estudiantes en las PAU sobre inferencia estadística.

Para lograr este objetivo, partiendo de la información ofrecida por la síntesis de las investigaciones previas y del análisis realizado para cumplir el objetivo específico anterior, se identifican las variables que serán objeto de estudio. Posteriormente, se realiza un análisis de contenido de los problemas, para identificar las categorías de cada una de las variables y, seguidamente, se analizan los datos para presentar la distribución de cada una de las variables globalmente y por año.

OE2.3. Analizar la idoneidad didáctica de los problemas de inferencia estadística propuestos en las PAU.

Se pretende realizar un análisis de algunos componentes de la idoneidad didáctica, siguiendo los indicadores de Godino (2009, 2013), para presentar una valoración cualitativa de las pruebas, teniendo en cuenta dichos componentes y los análisis realizados en el Capítulo 2.

3.3. HIPÓTESIS O EXPECTATIVAS INICIALES

Seguidamente, se formulan las hipótesis de investigación previstas en este estudio, que se discutirán en las conclusiones:

H2.1. Se espera encontrar una alta importancia de la inferencia estadística en las PAU, reflejada a través de un elevado número de problemas relacionados con la inferencia estadística.

El análisis semiótico de los problemas de inferencia propuestos en las PAU y su comparación con el contenido curricular de estadística en los decretos curriculares de Bachillerato mostrará el peso relativo que se da a la inferencia estadística dentro de estas pruebas. De acuerdo al análisis preliminar publicado por Díaz et al. (2014), así como de los resultados obtenidos en López-Martín, (2016), se espera encontrar un peso relativo mayor en las pruebas respecto a la proporción del contenido curricular en Bachillerato a este tema.

H2.2. Además, se observará un buen ajuste entre los contenidos de inferencia estadística fijados en el currículo y los considerados en las pruebas.

El análisis semiótico de la solución de los problemas permitirá identificar los objetos matemáticos sometidos a evaluación. De acuerdo al análisis preliminar publicado por Díaz et al. (2014), así como en los resultados obtenidos en López-Martín, (2016), la comparación de estas configuraciones de objetos con los contenidos y estándares curriculares mostrará un buen ajuste general.

H2.3. Se espera encontrar una discrepancia entre el contenido incluido en el currículo en relación al contraste de hipótesis y el evaluado en las PAU, con una mayor parte de problemas en las PAU centrados en la inferencia sobre la media de la población.

Aunque en el currículo que han cursado los estudiantes en los pasados años (MEC, 2007) se sugiere estudiar la inferencia estadística tanto para una como para dos poblaciones y en relación a la media y proporción, esperamos que los problemas incluidos en las PAU se centren en la inferencia sobre una media. La menor dificultad de estos problemas, en comparación con otros citados en los diseños curriculares, como los de comparación de medias de distribuciones normales con desviación típica conocida o incluso que los problemas de inferencia sobre una proporción, puesto que, en este caso, los estudiantes deben aplicar la corrección de continuidad, puede ocasionar la sobrerrepresentación de este tipo de problemas.

H2.4. Al estudiar el contexto de los problemas, se espera encontrar una proporción elevada de problemas descontextualizados.

De acuerdo al estudio realizado por Carretero (2014) sobre los problemas de probabilidad y en el trabajo de Alvarado (2007) sobre libros de texto, esperamos que haya una alta proporción de ejercicios descontextualizados en los problemas de las PAU analizados.

3.4. METODOLOGÍA

3.4.1. Descripción general de la metodología

La metodología empleada pretende dar respuesta a problemas concretos mediante la selección intencionada de una muestra (Hernández y Carpio, 2019). Por consiguiente, el objetivo de esta investigación es de tipo descriptivo, no siendo la intención extrapolar las conclusiones a otras pruebas diferentes a las analizadas. Los resultados obtenidos permitirán aportar resultados útiles para los especialistas universitarios que forman parte de las Comisiones Coordinadoras de las PAU.

Se ha efectuado un estudio longitudinal de análisis de las pruebas propuestas en las PAU de la modalidad del Bachillerato de Humanidades y Ciencias Sociales en la Comunidad Autónoma Andaluza, desde el curso académico 2002-2003 hasta el curso 2013-2014, en la asignatura de Matemáticas Aplicadas a las Ciencias Sociales II. Los contenidos de esta materia son cursados en segundo curso de Bachillerato en la Modalidad de Humanidades y Ciencias Sociales, y es la continuación de Matemáticas

Aplicadas a las Ciencias Sociales I abordada en el primer curso de Bachillerato de la mencionada modalidad.

De acuerdo con Díaz-Barriga y Luna-Miranda (2014), el proceso de investigación que se lleva a cabo se enmarca dentro de la tipología inductiva, típico de la metodología cualitativa de tipo documental, donde se pretende analizar los objetos matemáticos implícitos en la descripción de contenidos, criterios de evaluación y estándares de aprendizaje descritos en las pruebas de acceso a la universidad.

Se utiliza el análisis de contenido, que, de acuerdo a Krippendorff (2013), es una técnica de investigación destinada a formular, a partir de ciertos datos, inferencias reproducibles y válidas que puedan aplicarse a su contexto. En la misma línea, Weber (1990, p. 9) indica que “el análisis del contenido es un método de la investigación que utiliza un sistema de procedimientos para hacer inferencias válidas del texto”. El empleo de esta técnica permitirá, mediante la división de un texto en unidades, clasificar un número reducido de categorías en función de variables subyacentes, para posteriormente realizar inferencias sobre su contenido (Krippendorff, 2013). El plan de actuación a seguir será (Bisquerra, 2004):

1. Recolección de las PAU en la Comunidad Autónoma de Andalucía en el periodo 2003-2014 de la materia de Matemáticas Aplicadas a las Ciencias Sociales II. La documentación se ha extraído de la página oficial de la Junta de Andalucía.
2. Resolución de los problemas de inferencia estadística para identificar el contenido estadístico y como consecuencia, determinar las variables.
3. Identificación de las categorías de las variables objeto de estudio y codificación para su posterior análisis.
4. Revisión de la codificación por parte de varios investigadores para asegurar la fiabilidad de la codificación de los casos en que había duda sobre la categoría a asignar en alguna variable y problema.
5. Análisis descriptivo de la frecuencia de las variables y la realización de gráficos estadísticos elementales, tanto globalmente como por año nos permite obtener

conclusiones sobre la presencia global de los mismos, su distribución a lo largo del tiempo y la dificultad de los problemas.

Los resultados principales consisten en ejemplos prototípicos de las diferentes categorías, junto a gráficos que resumen la distribución estadística de las variables en los problemas y permiten comparar las diferentes categorías entre sí y su evolución temporal.

3.4.2. Muestra de problemas analizados

Como se ha indicado, se han revisado las pruebas de acceso a la universidad en Andalucía de 12 años consecutivos, considerando tanto las pruebas aplicadas como las de reserva. La prueba correspondiente a Matemáticas Aplicadas a las Ciencias Sociales II consta de dos opciones (opción A y opción B), cada una de ellas formada por cuatro ejercicios y donde el estudiante elige una de ellas. Tanto la opción A, como la opción B, están divididas en cuatro ejercicios. El primer ejercicio pertenece al Bloque de contenidos de Números y Álgebra, el segundo al Bloque de Análisis, y los dos últimos corresponden al Bloque de Estadísticas y Probabilidad, concretamente, a Probabilidad e Inferencia Estadística. Hasta el curso 2008/2009 la calificación correspondiente al Bloque de Estadística y Probabilidad suponía el 40% de la calificación total, aumentando a un 50% a partir del siguiente curso académico, porcentaje que se mantiene en la actualidad.

En lo que se refiere al análisis que se desarrolla en el presente capítulo, dentro del Bloque de Estadística y Probabilidad, nos restringimos al cuarto ejercicio que corresponde a los ejercicios de contenidos relacionados con la inferencia estadística. De cada año se revisaron los seis modelos disponibles (de los seis modelos, por sorteo, se selecciona la prueba de junio y la de septiembre), con dos opciones, cada una de ellas. Dado que el periodo seleccionado corresponde a 12 años, se han analizado 144 pruebas de acceso a la universidad.

En una primera toma de contacto se ha comprobado que hay enunciados de problemas con varios apartados, que incluyen preguntas que abarcan más de un campo de problema. Por tal motivo, el análisis de esta variable se ha realizado teniendo en cuenta cada uno de los apartados independientes recogidos en los problemas. De esta

forma la muestra asciende a 267 apartados. A continuación, se analiza un ejemplo de problema propuesto en cada uno de los campos de problemas y su resolución.

El análisis realizado en la presente memoria de tesis está basado en el enfoque ontosemiótico (Godino, 2002; Godino et al., 2007; 2020) en el que los “objetos matemáticos” son determinados a partir de las respuestas a situaciones problemáticas extra o intra-matemáticas. Mediante un análisis semiótico (Godino et al., 2007) se ha analizado el contenido matemático, determinando los objetos matemáticos implícitos en dicha resolución.

3.4.3. Variables analizadas

El estudio parte de la clasificación de los contenidos recogidos en el Decreto de Educación Secundaria Obligatoria y Bachillerato (MEC, 2007), que fueron presentados en Capítulo 1. A continuación, se han resuelto los problemas con el fin de analizar su contenido matemático y, así, determinar las variables que serán objeto de estudio.

V1. Campo de problema considerado. Se considera esta variable debido a la importancia que tiene en el marco teórico tanto de los problemas como de las prácticas necesarias para resolverlos, pues estas últimas determinan el significado de un objeto matemático (Godino, 2002; 2017; Godino et al., 2007, 2020). En consecuencia, si las prácticas requeridas para resolver diferentes problemas son distintas, esto incidirá en la diferencia en el significado evaluado en cada uno de los problemas. Teniendo en cuenta estas consideraciones, se han diferenciado tres campos principales de problemas:

A) *Muestreo*; El propósito es diseñar una muestra elegida al azar y formada por elementos independientes, para realizar un proceso de inferencia. En este campo se contempla, por un lado, los diferentes métodos de selección (muestreo aleatorio simple, estratificado, por conglomerados, etc.) e identificación de la distribución muestral.

B) *Intervalos de confianza*; La finalidad es estimar un parámetro poblacional mediante un intervalo de confianza a partir de una muestra representativa extraída de dicha población e indicar una medida del error de estimación.

C) *Contraste de hipótesis*; Se trata de poner a prueba una hipótesis, generalmente sobre un parámetro poblacional, a partir de los datos obtenidos en una muestra de la población.

Dichos campos de problema corresponden a los principales contenidos curriculares de inferencia estadística en Bachillerato y la investigación didáctica describe dificultades específicas para cada uno de ellos, como se ha visto en el Capítulo 2. Los dos primeros se han subdividido a su vez en dos subcampos de problema, que serían los siguientes:

A1) Enumerar las diferentes *muestras de una población*, que corresponde a un problema de naturaleza combinatoria, aunque supone el conocimiento de los diferentes tipos de muestreo por parte de los estudiantes y *cálculo de estadísticos de las distribuciones muestrales en poblaciones finitas*.

A2) Identificar la *distribución muestral de un estadístico* o realizar *cálculo de probabilidades* con la misma; para ello el estudiante ha de diferenciar los tres tipos de distribución que aparecen en el proceso de muestreo (Harradine et al., 2011), recordar el teorema central del límite e identificar los parámetros de la distribución normal correspondiente.

Además, respecto al intervalo de confianza, también se consideran dos subcampos:

B1) *Calcular o interpretar un intervalo de confianza*; Se trata de obtener un intervalo de confianza para un parámetro, a partir de los datos de una muestra y/o interpretar el significado del intervalo obtenido.

B2) *Determinar la relación entre confianza, error y tamaño muestral*, con objeto de determinar alguno de los tres elementos a partir de la información conocida sobre el resto de ellos.

V2. Modelo probabilístico de la población. Los modelos de distribución, junto a la probabilidad y distintas técnicas descriptivas, constituyen un pilar fundamental que permite obtener información de una población no observada en su totalidad, a partir de los datos suministrados por una pequeña parcela de dicha población. La selección de

una variable condicionará un tipo concreto de distribución de probabilidad y, por tanto, supeditará las estimaciones y los contrastes de hipótesis.

En el caso de la distribución normal o distribución gaussiana, esta constituye la piedra angular en los métodos de estimación y formulación de contrastes de hipótesis. Dicha distribución además de ser aplicable a variables con distribución esencialmente normal, por ejemplo, la media muestral, también puede ser empleada con variables de distribuciones no normales o incluso puede ser usada para realizar estimaciones de parámetros tales como la proporción siempre y cuando se garanticen las condiciones necesarias para aplicar el teorema central del límite.

En el estudio detallado en la Sección 3.6 se presentan otros modelos probabilísticos como la distribución uniforme discreta y distribución binomial, recogidas en las pruebas.

V3. Parámetro poblacional que debe ser estimado. Partiendo del supuesto de que la distribución de la población en estudio es conocida, la inferencia paramétrica se centra en realizar inferencias sobre los parámetros desconocidos del modelo probabilístico de la población.

En base a los contenidos recogidos en el Real Decreto de Educación Secundaria y Bachillerato (MEC, 2007) y a nuestro análisis preliminar de los problemas, los parámetros poblacionales a estimar o a los que se desea construir la distribución muestral que se detallan en la Sección 3.7 son: la media, varianza y proporción. Es importante que el estudiante diferencie estadístico y parámetro, punto que fue confuso en investigaciones como las de Vallecillos (1994).

V4. Contexto del problema. El contexto de donde se tomaron los datos tiene en estadística una especial importancia con el objeto de transmitir una estadística con sentido (Moore, 2010).

Las pruebas de evaluación *Programme for International Student Assessment*, conocidas como pruebas PISA, están orientadas a evaluar la competencia matemática. Concretamente, evalúa “la capacidad de formular, emplear e interpretar cuestiones matemáticas en diferente tipo de contextos” mediante la resolución de tareas relacionadas con la vida real (MCD, 2013, p. 11). En dichas pruebas participan

estudiantes de 15 años de 65 países, incluido España; por lo tanto, sería deseable que los contextos sugeridos en PISA también se tengan en cuenta en las PAU.

Los contextos considerados en el presente estudio, y que son detallados en la Sección 3.8, son: personal, profesional, social y científico. Aquellos problemas que no se han enmarcado en alguno de los contextos mencionados, han sido clasificados como problemas sin contexto o problemas descontextualizados.

3.5. CAMPOS DE PROBLEMAS, PRÁCTICAS MATEMÁTICAS Y CONFIGURACIÓN DE OBJETOS MATEMÁTICOS ASOCIADOS

El campo de problemas es un elemento fundamental en el enfoque ontosemiótico (EOS) (Godino y Batanero, 1994; 1998; Godino et al., 2007), pues determina las prácticas y objetos matemáticos involucrados en la actividad matemática y, por tanto, los objetos matemáticos que emergen de las mismas. Para estos autores, una situación-problema es cualquier aplicación extra-matemática o matemática que índice la actividad de matematización. En el caso de la enseñanza, podemos considerar que para el estudiante son problemas las tareas, ejercicios o actividades planteadas que inducen una actividad matemática.

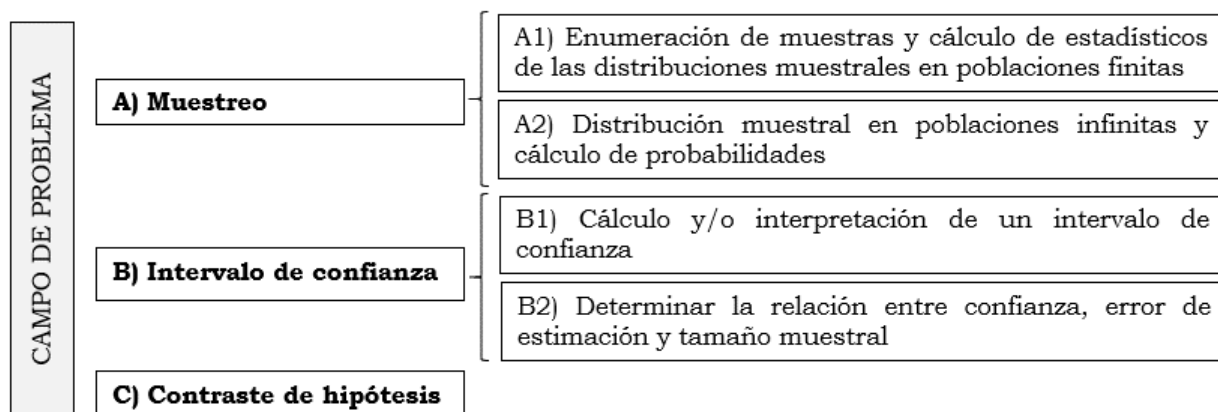
En el EOS no interesa cada problema particular, sino los tipos o campos de problemas que reúnen características similares. También en la enseñanza los problemas propuestos a los estudiantes suelen estar agrupados en tipos o campos de problemas. Puesto que el significado del objeto matemático se define a partir del conjunto de prácticas asociadas a los campos de problemas típicos, los campos de problemas que se propongan en las PAU determinarán el significado institucional evaluado en las mismas. Indirectamente, esto influirá en el significado institucional pretendido en la enseñanza de la inferencia en este nivel educativo y en el significado personal construido por los estudiantes que se preparan para abordarlas. La identificación de los campos de problemas también se considera como parte del análisis de los campos de actuación tal y como pone de manifiesto Rico (2006a).

En base a estas consideraciones, se ha realizado la clasificación de cada uno de los distintos tipos de problemas de inferencia estadística de las PAU en la muestra analizada. Para tal fin, se realizó la lectura de los enunciados de cada problema y

posteriormente se resolvieron, agrupando aquellos que eran más similares para determinar las categorías que hemos utilizado en esta variable. A modo de resumen, la Figura 3.5.1 recoge los campos de problemas identificados en el análisis de contenido de los problemas.

Dada la importancia que tiene esta variable, a lo largo de la sección, y con el fin de realizar el análisis semiótico del contenido evaluado, se resuelve un ejemplo típico en cada campo, identificando en la solución las prácticas matemáticas requeridas y la configuración de objetos matemáticos asociada. Posteriormente, se realiza una síntesis del contenido evaluado en el conjunto de campos de problema y se analiza el grado de relación que existe entre el significado institucional pretendido y el significado institucional evaluado para la inferencia en el citado nivel educativo.

Figura 3.5.1. Campos de Problemas Identificados en las PAU



3.5.1. Análisis semiótico de problemas propuestos

A lo largo de este apartado se analizan con detalle los campos de problemas (véase Figura 3.5.1). De los 144 problemas que forman la muestra, se ha seleccionado un ejemplo correspondiente a cada una de las categorías. Con el fin de analizar la diversidad y especificidad de objetos matemáticos que contiene la resolución del ejemplo, se ha resuelto minuciosamente. Con ello se muestra que la complejidad de los diferentes campos de problemas no es equivalente.

Sobre dichos problemas se realiza un análisis semiótico. Este análisis es propuesto por Godino (2002) para tener en cuenta la diversidad de objetos puestos en juego en la actividad matemática, multitud de actos y procesos de interpretación y el papel de los contextos en dichos procesos semióticos. La noción de función semiótica permite interpretar el conocimiento y la comprensión como conjunto de funciones semióticas que se pueden establecer en relación a un cierto objeto. Aunque pudiera pensarse que el interpretar los objetos matemáticos como sistemas de prácticas implica un relativismo en la visión de los objetos matemáticos, el dilema se resuelve al aceptar que el matemático identifica la misma estructura formal en una variedad de prácticas y objetos emergentes de las mismas (Wilhelmi, Godino y Lacasta, 2007).

El análisis semiótico consiste en la investigación sistemática de los significados (contenidos de las funciones semióticas) que se ponen en juego en una actividad matemática, por ejemplo, la resolución de un problema o una secuencia de instrucción (Godino, 2002). El resultado del análisis se puede comparar con el significado de referencia para detectar diferencias y prever conflictos semióticos potenciales.

A continuación, aplicamos este tipo de análisis a la solución prevista en cada uno de los campos de problemas identificados en las pruebas PAU.

A1. Enumeración de muestras y cálculo de estadísticos de las distribuciones muestrales en poblaciones finitas

El objetivo principal de la estadística inferencial es la obtención de conclusiones acerca de una población no observada a partir de una muestra, que se supone es representativa del total de elementos que compone la población. La selección de una muestra aleatoria formada por elementos independientes de la población con un número suficiente de elementos es un proceso que adquiere gran importancia en el campo de la inferencia estadística con objeto de proporcionar seguridad estadística en los resultados (Chance et al., 2004).

Con objeto de llevar a cabo las generalizaciones deseadas a partir de la muestra seleccionada, se calculan algunos resúmenes estadísticos (media muestral, varianza muestral, desviación típica muestral, proporción muestral, etc.), los cuales permiten estimar el valor o valores de los parámetros poblacionales (en nuestro caso: media, varianza y proporción poblacional).

Dentro de este bloque, nos centramos en estudiar aquellos problema-enunciados que incluyen los contenidos mencionados anteriormente. Concretamente, en este apartado se ha analizado los problemas que tratan los siguientes temas:

- Enumeración de las muestras que pueden formarse a partir de una población finita con un tipo particular de muestreo; es decir, producir un listado de todas las muestras posibles en dichas condiciones, lo que sería un problema combinatorio.
- Cálculos de los resúmenes estadísticos de una o varias muestras; problema que, en realidad quedaría dentro de la estadística descriptiva, aunque implica que el estudiante debe diferenciar entre estadístico y parámetro.

Generalmente, los datos que recoge el enunciado de un problema catalogado dentro de este campo son: el tipo de población de partida, es decir la variable que se estudia, la distribución de dicha variable y los parámetros que la determinan; el tamaño de la muestra y las condiciones del muestreo aleatorio (que puede ser muestreo aleatorio simple, muestreo sistemático y muestreo estratificado). Como ejemplo de este campo de problema, se ha seleccionado el problema 3B (prueba de reserva) propuesto en el curso académico 2013-2014, cuya solución prevista se analiza a continuación.

Problema 3B (2014)

- a) Determine todas las muestras de tamaño 2 que, mediante muestreo aleatorio simple, se pueden extraer del conjunto $\{6, 9, 12\}$ y calcule la varianza de las medias de estas muestras.
- b) Una empresa fabrica cuatro productos A, B, C y D, de los que elabora diariamente 40, 15, 25 y 120 unidades respectivamente. Si un día se quiere elaborar una muestra de 40 unidades con los productos fabricados, por muestreo aleatorio estratificado con afijación proporcional, ¿qué número de unidades de cada producto se debe elegir?

Solución prevista al apartado a)

Las prácticas matemáticas requeridas para resolver este problema son las siguientes:

En el primer apartado de este problema, el estudiante debe interpretar el enunciado para asumir que parte de una población finita discreta de distribución uniforme (con tres elementos equiprobables), puesto que puede aplicarse el principio de indiferencia, y se puede asignar probabilidad a estos elementos utilizando la concepción clásica de la misma (Batanero y Borovcnik, 2016). A continuación, se debe extraer aleatoriamente todas las muestras de un tamaño determinado (dos elementos) y

mediante una enumeración sistemática, producir un listado de las mismas. Posteriormente, debe calcular para cada una de las muestras el estadístico de interés (la varianza muestral). Por tanto, además de recordar los conceptos de muestra, muestreo aleatorio simple y varianza, requiere un razonamiento combinatorio para describir todas las muestras.

En relación a la resolución del primer apartado es necesario comprender el tipo de muestreo que se está realizando, pues los resultados obtenidos mediante el muestreo aleatorio con o sin reemplazamiento no coinciden entre sí (las distintas muestras y el número total no coinciden por ambos métodos) (Moore, 2010). Dado que en el enunciado del problema no se especifica lo contrario, se supone que el muestreo es con reemplazamiento, es decir, es posible extraer una muestra en la que los dos elementos que la forman son iguales.

Partiendo de esta premisa, se realiza la enumeración de todas las posibles muestras de dos elementos, obteniendo como resultado en total nueve muestras diferentes: (6, 6), (6, 9), (6, 12), (9, 6), (9, 9), (9, 12), (12, 6), (12, 9) y (12, 12). Se observa la necesidad de tener en cuenta el orden, pues la muestra formada por dos elementos diferentes tiene doble probabilidad que cuando los elementos son iguales. Este es un punto de potencial dificultad, pues los estudiantes suelen cometer el llamado *error de orden* (no tener en cuenta el orden cuando es requerido) (Batanero et al., 1994).

Para obtener la varianza de cada una de las medias muestrales, se requiere implícitamente el concepto de distribución muestral de la media, el cual se define a través de las medias de cada una de las muestras. Es decir, el estudiante debe concebir la media muestral como una variable aleatoria con su distribución, punto que es difícil según Schuyten (1991). Partiendo de dicha distribución muestral, la varianza se obtiene, a partir de la expresión $\sigma_{\bar{x}}^2 = \frac{\sum(\bar{x}_i - \mu_{\bar{x}})^2 n_i}{n}$ o equivalentemente $\sigma_{\bar{x}}^2 = \frac{\sum \bar{x}_i^2 n_i}{n} - (\mu_{\bar{x}})^2$ siendo \bar{x}_i los valores de las medias de cada muestra y n_i la frecuencia absoluta de los mismos. Un posible conflicto en este cálculo es que el estudiante confunda estadístico con parámetro, concretamente la varianza poblacional con la varianza de la distribución muestral o con la varianza de los datos, error observado en investigaciones como las de Cañadas et al. (2012) o Ramos et al. (2009).

Para facilitar el cálculo, además de la calculadora, se puede recurrir al uso de una tabla estadística, véase Tabla 3.5.1. Haciendo uso de la información recogida en dicha tabla, se tiene que la media y la varianza de la distribución muestral son respectivamente $\mu_{\bar{x}} = \frac{\sum \bar{x}_i n_i}{n} = \frac{81}{9} = 9$ y $\sigma_{\bar{x}}^2 = \frac{756}{9} - (9)^2 = 3$.

De forma alternativa, y en base al teorema central del límite, es posible obtener la solución del primer apartado. Dicho teorema establece una relación entre la media y desviación de una población no especificada, y la distribución de las medias de las muestras aleatorias con elementos independientes de tamaño n procedentes de dicha población. Concretamente, el teorema central del límite permite saber que la variable aleatoria media muestral, \bar{x} , tiene una distribución con media igual a la media poblacional (μ) y varianza igual a la varianza poblacional dividida por el tamaño muestral (σ^2/n) (Batanero y Borovcnik, 2016). Por tanto, en este caso, dado que la media y la varianza de poblacional son respectivamente $\mu = \frac{6+9+12}{3} = 9$ y $\sigma^2 = \frac{6^2+9^2+12^2}{3} - 9^2 = 6$, entonces se verifica que $\mu_{\bar{x}} = 9$ y $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{6}{2} = 3$, coincidiendo con los resultados obtenidos anteriormente. Se destaca el hecho de que el desarrollo alternativo genera mayor complejidad a los estudiantes por las dificultades que esto tienen en relación a la comprensión del dicho teorema (Alvarado, 2007).

Tabla 3.5.1.

Organización de Cálculos en la Resolución del Problema 3B (2014)

Medias muestrales (\bar{x}_i)	n_i	$\bar{x}_i \times n_i$	$\bar{x}_i^2 \times n_i$
6	1	6	36
7,5	2	15	112,5
9	3	27	243
10,5	2	21	220,5
12	1	12	144
Sumas	9	81	756

Solución prevista al apartado b)

En este apartado se pide determinar la composición de una muestra de tamaño 40, mediante un muestreo aleatorio estratificado con afijación proporcional, facilitando como información el tamaño total de cada estrato.

Con respecto a la resolución es necesario conocer y distinguir los diferentes tipos de muestreo aleatorio que se abordan en los contenidos de la asignatura (muestreo aleatorio simple, estratificado y sistemático). Al tratarse de un muestreo aleatorio estratificado con afijación proporcional, existe una relación de proporcionalidad entre el número total de elementos de todos los estratos ($N=N_1+N_2+\dots+N_k$, donde N_i es el tamaño del estrato i) y el número total de elementos de la muestra extraída ($n=n_1+n_2+\dots+n_k$, siendo n_i el tamaño de la muestra i correspondiente al estrato i), es decir, se verifica que $\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_k}{N_k} = \frac{n}{N}$. En este caso, se establece la relación siguiente:

$$\frac{n_1}{40} = \frac{n_2}{15} = \frac{n_3}{25} = \frac{n_4}{120} = \frac{40}{40 + 15 + 25 + 120} = \frac{40}{200}$$

a partir de la cual se tiene que $\frac{n_1}{40} = \frac{40}{200}$ y, como consecuencia la muestra correspondiente al estrato primero estaría compuesta por 8 unidades, es decir, $n_1 = 8$. Procediendo de forma análoga se obtienen el número de unidades para cada estrato, $n_2 = 3$, $n_3 = 5$ y $n_4 = 24$.

Los contenidos implicados en la tarea analizada son esenciales para el razonamiento posterior sobre la realización de inferencias respecto a las características de una población cuyo tamaño excesivamente elevado provoca una imposibilidad material y temporal de muestrear todos sus individuos. Dado que en la práctica se recurre a una muestra para inferir datos de esa población, es necesario garantizar la representatividad de la muestra prestando atención al tamaño de esta y al proceso de selección de los elementos que la forman.

Del análisis semiótico realizado, se presenta en la Tabla 3.5.2 la configuración epistémica ligada a la resolución del problema. Se observa que claramente nos encontramos dentro de la inferencia frecuencial, al suponer el muestreo repetido y analizar la distribución muestral del estadístico. Además, se considera el parámetro como constante y el estadístico como variable aleatoria. Los principales procedimientos aplicados son combinatorios (enumeración sistemática) y cálculo de estadísticos, que podría simplificarse con la calculadora. En este campo de problemas no surge todavía la necesidad de leer las tablas de la distribución normal, tipificar o resolver desigualdades.

A2. Distribución muestral en poblaciones infinitas y cálculo de probabilidades

Dentro de esta categoría de campos de problemas, se han considerado todos aquellos en los que se pide describir la distribución muestral de un cierto estadístico en poblaciones infinitas y el cálculo de probabilidades a partir de dicha distribución muestral, utilizando algún modelo teórico de distribución. Se parte de la premisa que la selección de la muestra se ha realizado mediante un muestreo aleatorio con reemplazamiento, implicando la obtención de un número infinito de muestras aleatorias independientes de tamaño n proveniente de la población de interés.

Tabla 3.5.2.

Configuración Epistémica Asociada al Campo de Problemas A1

Conceptos	Población y muestra, elementos Parámetro y estadístico Muestreo aleatorio simple Muestreo estratificado, afijación proporcional Muestreo con y sin reemplazamiento: diferencias Distribución muestral Probabilidad, asignación clásica; principio de indiferencia Modelo teórico de distribución Distribución uniforme discreta
Propiedades	La muestra es una parte de la población Representatividad muestral Variabilidad muestral El parámetro es fijo y constante El estadístico es una variable aleatoria Relación entre los estadísticos de la población y de la distribución muestral Relación de la variabilidad muestral con el tamaño de la muestra
Procedimientos	Enumeración de las muestras de una población finita Formar muestras de una población finita Cálculo de los estadísticos en las muestras Cálculo de los estadísticos de la distribución muestral Cálculo del tamaño de la muestra usando proporciones Uso de la calculadora
Lenguaje	Verbal y simbólico
Argumento	Deductivo

Cuando se recogen diferentes muestras de la misma población, se verifica que el estadístico puede tomar diferentes valores, dependiendo de los elementos que determinan la muestra. Por ejemplo, de una misma población se podrían extraer diversas muestras de mismo tamaño, pero con medias muestrales distintas. Este hecho, provoca que el estadístico sea considerado como una variable aleatoria con propia

distribución, denominada *distribución muestral* (Moore, 2010). Así pues, la tarea que se pide al estudiante en este segundo campo de problemas es reconocer la distribución de probabilidad de las variables aleatorias asociadas a estadísticos muestrales procedentes de poblaciones infinitas. Dicha determinación permitirá realizar inferencia respecto a los parámetros desconocidos de una población. Además, el conocer la distribución muestral de un estadístico nos permitirá hallar probabilidades asociadas al suceso consistente en que el estadístico muestral tome ciertos valores (Behar, 2001).

Los resúmenes estudiados en Bachillerato son la media muestral y la proporción muestral, que son parámetros correspondientes a las distribuciones normal y binomial respectivamente. En el caso de una distribución normal de desviación típica conocida, la media de la muestra se distribuye también normalmente, debido a las propiedades reproductivas de esta distribución (Moore, 2010). En caso de que la muestra tenga un tamaño suficiente ($n > 30$) se puede aplicar el teorema central del límite para deducir que, sea cual sea la población de partida, la distribución también es normal (Moore, 2010). Otras distribuciones muestrales implican a la distribución t de Student, cuando no se conoce la desviación típica de la población, pero no suelen estudiarse en el Bachillerato. Para el caso de la distribución binomial, aplicando también el teorema central del límite, la distribución de la proporción muestral se aproxima a la distribución normal en el caso de muestras de tamaños elevados, aunque en este caso se debe aplicar la corrección de continuidad.

Generalmente, en el enunciado se facilita la información referente al tamaño muestral, tipo de población y estadístico de interés. En el caso particular de analizar la media muestral, se informa explícitamente de la media y varianza poblacional, mientras que, si se trata de la proporción, el estudiante ha de recordar su fórmula (cociente entre el número de veces que se presenta un dato, n_i , respecto al total de datos que componen la muestra, n , es decir, n_i/n). Una vez determinada la distribución del estadístico en estudio, se suele pedir calcular probabilidades, como en el ejemplo que se analiza el problema 5B propuesto en la convocatoria de septiembre en el curso académico 2010-2011, que se resuelve a continuación.

Problema 5B (2011)

Sea X una variable aleatoria Normal de media 50 y desviación típica 4. Se toman muestras de tamaño 16.

- ¿Cuál es la distribución de la media muestral?
- ¿Cuál es la probabilidad de que la media muestral esté comprendida entre 47,5 y 52,5?

Para resolver el problema, se deben realizar el siguiente conjunto de prácticas matemáticas en cada uno de los dos apartados.

Solución prevista al apartado a)

Se debe recordar que, puesto que la distribución de partida es normal, debido a las propiedades reproductivas de dicha distribución, la distribución muestral de la media será también normal. Igual ocurriría, aunque de forma aproximada en base al teorema central del límite, si el tamaño de la muestra es mayor que 30 elementos, independientemente de la forma de la distribución en la población.

Aplicando la linealidad de la media, así como la linealidad de la varianza cuando las variables que se suman son independientes, se obtiene, de forma inmediata, la resolución del primer apartado, ya que si una variable aleatoria X se distribuye según un modelo probabilístico normal con media μ y desviación típica σ , entonces, en las condiciones especificadas en el párrafo anterior, se verifica que la variable aleatoria media muestral, \bar{X} , sigue una distribución $N\left(\mu, \sigma/\sqrt{n}\right)$. Por tanto, la resolución del apartado a) consiste en indicar que $\bar{X} \rightarrow N\left(50, 4/\sqrt{16}\right) = N(50, 1)$ y como consecuencia el estudiante ha de recordar las propiedades mencionadas.

Solución prevista al apartado b)

El hecho de describir la distribución de la media de las muestras permite obtener probabilidades sobre dicha variable, aplicando los procedimientos de tipificación y lectura de tablas de probabilidades que se conocen. Así pues, tipificando la distribución de la media muestral ($Z = (\bar{X} - 50)/1$), se obtiene la distribución normal tipificada, $N(0,1)$, y el estudiante puede consultar las tablas de probabilidades para la resolución del problema (Moore, 2010). A partir de estas tablas y aplicando las reglas simples de cálculo de probabilidades se obtiene:

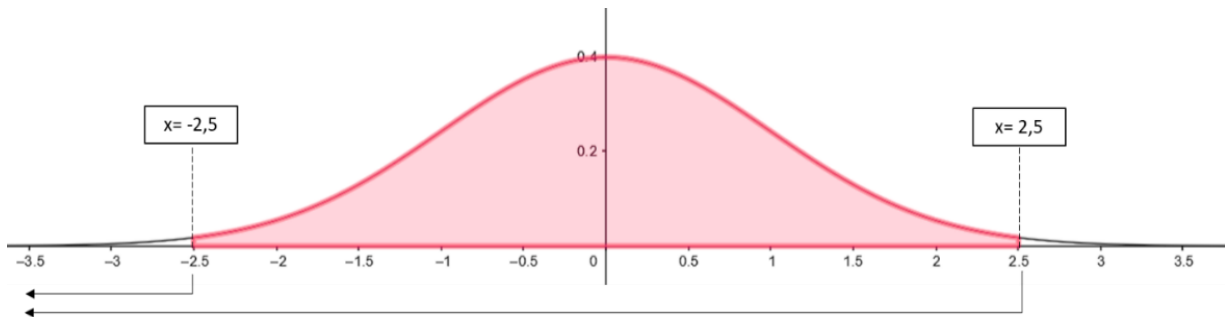
$$P\{47,5 \leq \bar{X} \leq 52,5\} = P\left\{\frac{47,5 - 50}{1} \leq \frac{\bar{X} - 50}{1} \leq \frac{52,5 - 50}{1}\right\} = P\{-2,5 \leq Z \leq 2,5\}$$

El estudiante ha de manejar correctamente las desigualdades para obtener los nuevos límites tipificados. Puesto que la tabla de la distribución normal que se emplea en Bachillerato sólo contiene los valores de probabilidad acumulada hasta un Z dado, el estudiante ha de usar las propiedades aditivas de la probabilidad y de simetría de la curva normal. A partir de la Figura 3.5.2, se tiene que la $P\{-2,5 \leq Z \leq 2,5\}$ (zona coloreada en Figura 3.5.2) se obtiene como:

$$P\{-2,5 \leq Z \leq 2,5\} = P\{Z \leq 2,5\} - P\{Z < -2,5\} = 0,9876.$$

Por lo que, la probabilidad de que la media muestral esté comprendida entre 47,5 y 52,5 es 0,9876, es decir, en el 98,76 % de las muestras seleccionadas de tamaño 16, la media estará en el intervalo delimitado por ambos valores.

Figura 3.5.2. Área Delimitada por la Curva de la Distribución Normal Estandarizada, $N(0,1)$, y las Rectas $x = -2,5$ y $x = 2,5$



Con el fin de identificar correctamente la distribución de la variable en estudio, el estudiante debe tener en cuenta la relación existente entre la media y varianza de la población y la media y varianza de las muestras de un tamaño determinado. Además, debe ser capaz de tipificar, aplicar las reglas básicas de cálculo de probabilidades en un intervalo según una variable aleatoria continua, emplear las reglas de simetría de la curva normal y leer correctamente las tablas de probabilidades de una normal tipificada, para conseguir así obtener la probabilidad pedida. En ambos tipos de problemas, el estudiante podría cometer errores de tipificación o de manejo de las inecuaciones

requeridas en el cálculo de probabilidades o bien de determinación de percentiles en la curva normal (Batanero et al., 2001).

Los problemas de cálculo de probabilidades tienen un punto de dificultad añadido cuando la variable en estudio es la proporción muestral. De acuerdo a Alvarado y Batanero (2007) el uso de la corrección de continuidad, requerida para el trabajo de probabilidades de una binomial mediante la aproximación a la normal puede generar cierta confusión. Esta dificultad ha provocado que algunos de los libros de texto dirigidos a los estudiantes de segundo de Bachillerato recomiendan llevar a cabo este tipo de ejercicios mediante un planteamiento incorrecto, sin considerar la corrección de continuidad, con el fin de simplificar el proceso (Colera et al., 2003).

Como consecuencia del conjunto de prácticas matemáticas descritas, el análisis semiótico permite identificar la configuración epistémica asociada al campo de problemas A2 (véase Tabla 3.5.3) donde se encuentran diferencias con el anterior campo de problemas y que se comentarán posteriormente.

B1. Cálculo o interpretación de un intervalo de confianza

Se contempla los problemas en los que se pide explícitamente la construcción de un intervalo de confianza para un cierto estadístico (generalmente la media o la proporción muestral) y los relacionados con la interpretación de la información aportada por el intervalo. La construcción de un intervalo de confianza depende de:

1. Las características de la población, es decir, del modelo probabilístico de la población (binomial, normal, etc.);
2. Los parámetros poblacionales desconocidos, como pueden ser por ejemplo la media, varianza y proporción poblacional;
3. Los valores de los estadísticos muestrales conocidos, como por ejemplo la media, varianza y proporción muestral;
4. El tamaño muestral;
5. El coeficiente de confianza o nivel de confianza.

Aunque a la hora de determinar un intervalo hay que tener en cuenta diversos factores y circunstancias, la forma de actuar en su construcción permanece invariable. En lo que sigue, se analiza el problema 1A (prueba de reserva del curso 2005-2006) en el que el parámetro a estimar es la proporción poblacional.

Tabla 3.5.3.*Configuración Epistémica Asociada al Campo de Problema A2*

Tipo	Objeto matemático
Conceptos	Población y muestra, elementos Parámetro y estadístico Muestreo aleatorio simple Muestreo con y sin reemplazamiento: diferencias Distribución muestral Probabilidad, asignación clásica; principio de indiferencia Modelo teórico de distribución Distribución normal Distribución normal tipificada Teorema central del límite
Propiedades	La muestra es una parte de la población Representatividad muestral Variabilidad muestral El parámetro es fijo y constante El estadístico es una variable aleatoria Linealidad de la media de la suma de variables aleatorias Linealidad de la varianza de la suma de variables aleatorias independientes Relación entre los estadísticos de la población y de la distribución muestral Relación de la variabilidad muestral con el tamaño de la muestra Condiciones de aplicación del teorema central del límite Aproximación normal a la distribución binomial Simetría de la curva normal
Procedimientos	Tipificación Cálculo de probabilidades en la distribución muestral Lectura de las tablas de la $N(0,1)$ Manejo de desigualdades Aplicación de propiedades de la probabilidad Corrección de continuidad, para el caso de la distribución de la proporción Uso de la calculadora
Lenguaje	Verbal, simbólico, gráfico
Argumento	Deductivo

A continuación, se describen las prácticas matemáticas requeridas en la resolución de este problema.

Problema 1A (2006)

De 500 encuestados en una población, 350 se mostraron favorables a la retransmisión de debates televisivos en tiempos de elecciones. Calcule un intervalo de confianza, al 99,5%, para la proporción de personas favorables a estas retransmisiones.

Solución prevista

En el caso de variables binarias, la proporción de elementos que poseen cierta característica o la proporción de "éxitos" de una secuencia de n ensayos en la población, denotada por p , recibe el nombre de *proporción poblacional*. Si en una población dicotómica, se consideran todas las posibles muestras de tamaño n que pueden extraerse de la mencionada población, se verificará que en cada una de ellas habrá una proporción de elementos con esa característica. Puesto que dicha proporción varía en función de la muestra seleccionada, entonces, la proporción muestral, denotada por \hat{p} , adquiere el carácter de variable aleatoria discreta distribuida según una binomial de parámetros n y p , es decir, $\hat{p} \sim B(n, p)$.

Siempre y cuando se verifique que $n > 10$, $n \times p > 5$ y $n \times (1 - p) > 5$, el teorema de Moivre-Laplace establece que se puede realizar una aproximación bastante buena de la variable discreta binomial a una variable normal con media $n \times p$ y desviación típica $\sqrt{n \times p \times (1 - p)}$. Dado que la proporción muestral se define como $\hat{p} = \frac{X}{n}$ donde $X =$ "número de éxitos en una muestra de tamaño n " $\sim B(n, p)$ entonces la proporción muestral, $\hat{p} \sim N\left(\frac{np}{n}, \frac{\sqrt{np(1-p)}}{n}\right) = N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. Como consecuencia, al considerar el proceso de tipificación, se obtiene que $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0, 1)$.

Un elemento básico en la construcción del intervalo es el coeficiente de confianza o nivel de confianza $(1 - \alpha) \times 100\%$. Dado este coeficiente, el objetivo es hallar los límites de un intervalo de valores (e_1, e_2) , de tal forma que se verifique que la probabilidad de que el parámetro se encuentre entre ambos valores coincida con el nivel de significación, es decir, $P(e_1 < p < e_2) = 1 - \alpha$. Su proceso de cálculo comienza identificando un intervalo de valores de la distribución normal tipificada, centrado en el origen y tal que el área de cada cola sea $\frac{\alpha}{2}$, es decir,

$$P\{-k < Z < k\} = 1 - \alpha \quad (3.5.1)$$

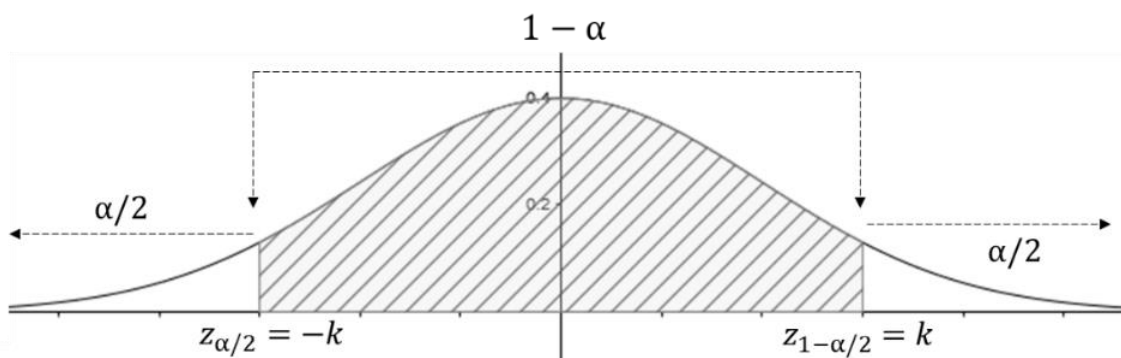
Realizando una lectura inversa de la tabla de la distribución normal se tiene que, los valores $-k$ y k que verifican la expresión (3.5.1) son aquellos que cumplen que

$P\{Z < -k\} = \alpha/2$ y $P\{Z < k\} = 1 - \alpha/2$, respectivamente (véase Figura 3.5.3). A partir de estos valores se despejan los extremos del intervalo de confianza de la variable Z .

En este caso, puesto que la proporción muestral se distribuye según una normal, tipificando, se obtiene que $P\left\{|\hat{p} - p| < k\sqrt{\frac{p(1-p)}{n}}\right\} = 1 - \alpha$ y como consecuencia:

$$P\left\{\hat{p} - k\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + k\sqrt{\frac{p(1-p)}{n}}\right\} = 1 - \alpha. \quad (3.5.2)$$

Figura 3.5.3. Selección de los Puntos Críticos (para una Distribución Normal Tipificada) en el Cálculo del Intervalo de Confianza



Una vez despejado el parámetro a estimar, se observa que los límites dependen del parámetro desconocido, véase expresión (3.5.2). Si n es suficientemente grande, se considera, como una solución satisfactoria, sustituir la proporción poblacional por su estimador, en este caso, la proporción muestral. Así pues, en base a los comentarios realizados, se obtiene que el intervalo de confianza de la proporción poblacional, a un nivel de confianza del $(1 - \alpha) \times 100\%$, viene dada por la expresión (3.5.3)²

$$\text{I.C.}_p = \left(\hat{p} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right). \quad (3.5.3)$$

² Nótese que $Z_{\frac{\alpha}{2}} = -Z_{1-\frac{\alpha}{2}}$

Haciendo uso de la tabla de la distribución normal tipificada se tiene que el cuantil 0,9975 corresponde al valor 2,81 y dado que la proporción muestral en nuestro caso es $\hat{p} = \frac{350}{500} = 0,7$ entonces se obtiene que el intervalo de confianza solicitado vendría dado por

$$\text{I.C.}_p = \left(0,7 - 2,81 \sqrt{\frac{0,7 \times 0,3}{500}}, 0,7 + 2,81 \sqrt{\frac{0,7 \times 0,3}{500}} \right) = (0,6421; 0,7576)$$

Como consecuencia del conjunto de prácticas matemáticas descritas, el análisis semiótico permite identificar la configuración epistémica asociada al campo de problemas B1 (véase Tabla 3.5.4), donde ahora se introducen nuevos conceptos, como el de intervalo de confianza y nivel de confianza y se utiliza la definición frecuencial de la probabilidad en su definición.

B2. Relación entre confianza, error de estimación y tamaño muestral

En la práctica, cuando se pretende obtener un intervalo de confianza para un parámetro poblacional siempre se intenta minimizar el error de estimación. En general, la expresión del intervalo de confianza depende del valor estimado del parámetro, obtenido a partir de la muestra, del tamaño muestral y del nivel de confianza seleccionado.

Tal y como se ha puesto de manifiesto anteriormente la selección de la muestra es un elemento que adquiere gran importancia ya que el tamaño de esta guarda una estrecha relación con el nivel de confianza y el error de estimación. Un aumento en el tamaño de la muestra implica tener una mayor precisión en la estimación y a la misma vez un menor error de estimación.

En algunos de los problemas propuestos en las PAU se pide al estudiante determinar uno de estos valores conocidos los otros, por ejemplo, determinar el tamaño de muestra para conseguir un intervalo de una cierta precisión considerando el mismo nivel de confianza o calcular el error de estimación que se está cometiendo, teniendo en cuenta el nivel de confianza establecido y el tamaño de la muestra. En pocas ocasiones se hallan problemas en los que se solicita el cálculo del nivel de confianza a partir de la información aportada por el resto de elementos.

Por todo lo expuesto, se ha visto recomendable analizar dentro de este bloque aquellos problemas en los que se solicita calcular el nivel de confianza, error de

estimación o tamaño muestral, después de haber fijado dos de ellos. Como datos proporcionados en el enunciado se suele identificar la población y el estadístico considerado o bien el intervalo de confianza. Un ejemplo de este tipo es el problema 3B incluido en la prueba de septiembre celebrada en el curso académico 2006-2007.

Tabla 3.5.4.

Configuración Epistémica Asociada al Campo de Problema B1

Tipos	Objetos matemáticos
Conceptos	Población y muestra, elementos, Parámetro y estadístico, Distribución muestral Muestreo aleatorio simple Probabilidad, asignación frecuencial Modelo teórico de distribución Distribución normal o binomial Distribución normal tipificada Teorema central del límite, Probabilidad condicional Intervalo de confianza Coeficiente de confianza o nivel de confianza; significado Error de estimación
Propiedades	La muestra es una parte de la población Representatividad y variabilidad muestral El parámetro es fijo y constante El estadístico es una variable aleatoria Linealidad de la media de la suma de variables aleatorias Linealidad de la varianza de la suma de variables aleatorias independientes Relación entre los estadísticos de la población y de la distribución muestral Relación de la variabilidad muestral con el tamaño de la muestra Condiciones de aplicación del Teorema Central del Límite Aproximación normal a la distribución binomial Los extremos del intervalo de confianza son variables No se sabe si el parámetro está comprendido en el intervalo El coeficiente de confianza da el % de intervalos que cubren el parámetro Relación entre amplitud de intervalo, confianza y tamaño muestral Relación entre nivel de significación y coeficiente de confianza
Procedimientos	Tipificación Cálculo de probabilidades en la distribución muestral Lectura de las tablas de la $N(0,1)$ Manejo de desigualdades Aplicación de propiedades de la probabilidad Corrección de continuidad, para el caso de la distribución de la proporción Cálculo de extremos del intervalo de confianza Uso de la calculadora
Lenguaje	Verbal y simbólico, gráfico
Argumento	Deductivo Inductivo empírico

A continuación, se describen las prácticas matemáticas requeridas en la resolución de los problemas de este campo.

Problema 3B (2007)

Se sabe que (45,13, 51,03) es un intervalo de confianza, al 95%, para la media de una variable aleatoria que sigue una distribución Normal con desviación típica 15.

a. ¿Cuál es el error cometido?

b. Calcule, con el mismo nivel de confianza, el tamaño muestral mínimo necesario para que el error no sea superior a 1,8.

Solución prevista al apartado a)

En este problema es necesario diferenciar entre estadístico (media de la muestra) y parámetro (media de la población) y comprender el significado del intervalo de confianza (rango de posibles valores para el parámetro). También hay que comprender el concepto de nivel de confianza (95% de los intervalos construidos de la misma población, con muestras del mismo tamaño, contendrán al parámetro; pero no sabemos si el intervalo particular lo contiene o no). Puesto que se da la distribución de la población como dato, y se trata de una distribución normal $N(\mu, \sigma)$, entonces se verifica que la variable aleatoria media muestral, denotada por \bar{X} , se distribuye según una distribución normal $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, donde n es el tamaño de la muestra (desconocida en el problema que se está analizando).

Para resolver el primer apartado, el estudiante ha de recordar, además de las propiedades anteriores y la definición de intervalo de confianza, que el error de estimación (error máximo admisible o cota de error de estimación) se define como la mitad de la amplitud del intervalo. Es decir, es la mitad de la diferencia existente entre el extremo superior e inferior del intervalo (véase expresión 3.5.4). En el caso particular que se está trabajando, el error de estimación es $E = \frac{1}{2} \times (51,03 - 45,13) = 2,95$.

$$E = \frac{1}{2} \times \left[\left(\bar{x} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \right] \quad (3.5.4)$$

Solución prevista al apartado b)

A diferencia del apartado a) en este caso se solicita el tamaño muestral necesario para garantizar que la diferencia existente entre la media muestral y la media poblacional

sea inferior a 1,8, manteniendo constante el resto de elementos. En este caso particular, haciendo uso de la expresión (3.5.5) se obtiene el tamaño muestral necesario que garantice el error de estimación deseado.

$$E = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}. \quad (3.5.5)$$

De la expresión (3.5.5) se deduce que a medida que aumenta el tamaño de la muestra la precisión en la estimación realizada será mejor, ya que el intervalo es más estrecho y el error de estimación será inferior.

Despejando n de la expresión (3.5.5) que $n = \left(Z_{1-\frac{\alpha}{2}} \frac{\sigma}{E} \right)^2 = \left(1,96 \frac{15}{1,8} \right)^2 = 266,7$, por tanto, se concluye que la muestra debe estar constituida al menos por 267 elementos. El resultado obtenido implicaría que, si repitiéramos el experimento, seleccionando una muestra de 267 elementos, en el 95% de las veces, se obtendría un error de estimación por debajo a 1,8, mientras que en el 5% restante obtendríamos un error superior.

A partir de la expresión del error de estimación, expresión (3.5.5), se pueden establecer las siguientes propiedades (Behar, 2001):

- a. A medida que aumenta el tamaño de la muestra (manteniendo constante el nivel de confianza) el error de estimación disminuirá;
- b. Manteniendo constante el tamaño de la muestra, el error aumentará conforme aumente el nivel de significación.

No obstante, y teniendo en cuenta las propiedades señaladas, si el objetivo es conseguir una mayor precisión en las estimaciones por medio de los intervalos de confianza, lo ideal es aumentar el tamaño muestral, ya que generalmente no es recomendable utilizar un nivel de confianza inferior al 90% ni del 100%. El análisis semiótico nos permite identificar la configuración epistémica asociada al campo de problemas B2 (véase Tabla 3.5.5), similares a las del problema B1.

C. Contraste de hipótesis

El último campo de problemas analizado ha sido el contraste de hipótesis. En este caso, se solicita al estudiante que compruebe si una afirmación estadística relativa al valor de un parámetro de la población es compatible con el resultado obtenido de la información aportada por una muestra aleatoria simple procedente de la misma. Las

diferencias que pueden presentarse es el tipo de población considerada (normal o binomial) y los apartados del problema o los datos aportados.

Tabla 3.5.5.

Configuración epistémica asociada al campo de problema B2

Tipo	Objeto matemático
Conceptos	Población y muestra, elementos, parámetro y estadístico Muestreo aleatorio simple Distribución muestral Probabilidad, asignación frecuencial, probabilidad condicional Modelo teórico de distribución Distribución normal o binomial, distribución normal tipificada Teorema Central del Límite Intervalo de confianza Coeficiente de confianza o nivel de confianza; significado Error de estimación, erro máximo admisible
Propiedades	La muestra es una parte de la población Representatividad y variabilidad muestral El parámetro es fijo y constante y el estadístico es una variable aleatoria Linealidad de la media de la suma de variables aleatorias Linealidad de la varianza de la suma de variables aleatorias independientes Relación entre los estadísticos de la población y de la distribución muestral Relación de la variabilidad muestral con el tamaño de la muestra Condiciones de aplicación del Teorema Central del Límite Aproximación normal a la distribución binomial Los extremos del intervalo de confianza son variables No se sabe si el parámetro está comprendido en el intervalo El coeficiente de confianza da el % de intervalos que cubren el parámetro Relación entre amplitud de intervalo, confianza y tamaño muestral El error de estimación es la mitad de la amplitud del intervalo de confianza
Procedimientos	Relación entre nivel de significación y coeficiente de confianza Tipificación Cálculo de probabilidades en la distribución muestral Lectura de las tablas de la N (0,1) Manejo de desigualdades Aplicación de propiedades de la probabilidad** Corrección de continuidad, para la distribución de la proporción Cálculo de extremos del intervalo de confianza Cálculo del tamaño mínimo de muestra para una precisión dada Uso de la calculadora
Lenguaje	Verbal y simbólico, gráfico
Argumento	Deductivo

Tomemos como ejemplo el siguiente, propuesto en la prueba de reserva de 2013. A continuación, se describen las prácticas matemáticas requeridas en la resolución de los problemas de este campo.

Problema 5A (2013):

Un director sanitario sostiene que el Índice de Masa Corporal (IMC) medio de los adolescentes de su distrito no supera el nivel 25 (sobrepeso). Para contrastar su afirmación toma una muestra aleatoria de 225 adolescentes que da como resultado un IMC medio de 26. Sabiendo que el IMC sigue una distribución Normal con desviación típica 5 discuta, mediante un contraste de hipótesis, si la afirmación del director sanitario es correcta, con un nivel de significación del 5%.

Solución prevista

El problema pide contrastar una hipótesis sobre la media de una población normal con media desconocida y desviación típica igual a 5, $N(\mu, 5)$. El procedimiento comienza estableciendo la hipótesis nula y la hipótesis alternativa. Recuérdese que la hipótesis alternativa sería la que interesa al investigador (Batanero, 2000) y la nula la contraria o de no efecto. En el caso del problema 5A se establecen las siguientes hipótesis:

$$H_0: \mu \leq 25; H_1: \mu > 25$$

Al analizar la hipótesis alternativa, se observa que se trata de un contraste unilateral derecho, pues la hipótesis alternativa indica que la media poblacional debe ser mayor a 25. Se pide tomar una decisión consistente en rechazar o no la hipótesis con un nivel de significación dado (5%). Para ello, deben determinarse los valores críticos que delimitaran las regiones de aceptación y rechazo. En la resolución del problema, será necesario recordar los siguientes conceptos y propiedades (Batanero, 2000; Vallecillos, 1994):

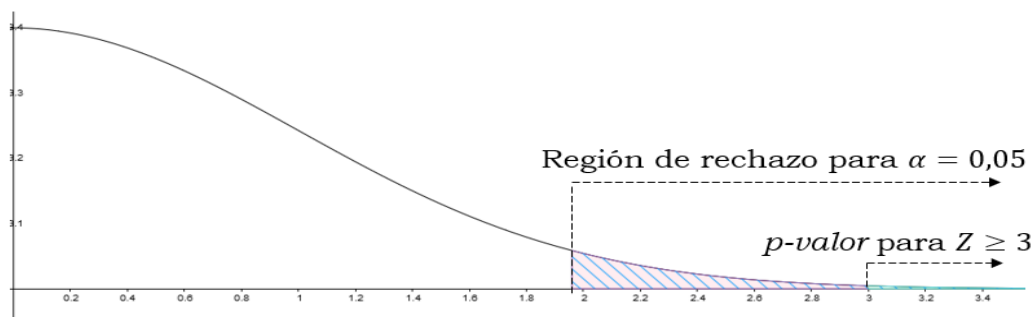
- En la elaboración de un contraste de hipótesis sobre la media de la población, se utiliza como estimador de μ el estimador insesgado la media muestral, cuyo valor debe ser identificando a partir de los datos que facilita el problema;
- Si la variable aleatoria X se distribuye según un modelo normal $N(\mu, \sigma)$, entonces se verifica que para muestras de tamaño 225 la variable aleatoria media muestral se distribuye según $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(\mu, \frac{5}{\sqrt{225}}\right) = N\left(\mu, \frac{1}{3}\right)$;

- Bajo la suposición de que la hipótesis nula es cierta, es decir, que $\mu = 25$, el estadístico de contraste estandarizado, $Z = \frac{\bar{x}-25}{1/3}$, es igual a 3 cuando la media muestral toma el valor 26.

En dicha distribución normal tipificada, el nivel mínimo de significación o p -valor, se obtendría como la probabilidad de obtener un valor superior o igual a 3, suponiendo la hipótesis nula H_0 cierta. En nuestro caso, $p = P(Z > 3 | H_0 \text{ cierta}) = 0,0013$. Así pues, si la hipótesis nula fuera cierta, es decir, si el índice medio de masa corporal en su distrito fuese menor o igual a 25, la probabilidad de obtener una muestra de 225 adolescentes con un índice medio igual o mayor a 26 sería 0,0013. Comparando la probabilidad de observar el valor dado (p -valor) con el nivel de significación prefijado de antemano, se rechazaría la hipótesis nula ya que $0,0013 < 0,05$ (véase Figura 3.5.4). Por lo tanto, lo razonable sería rechazar la afirmación del director sanitario ya que los datos empíricos no apoyan la hipótesis nula planteada.

Como consecuencia del conjunto de prácticas matemáticas descritas, el análisis semiótico nos permite identificar la configuración epistémica asociada al campo de problemas C (véase Tabla 3.5.6) donde aparecen los nuevos elementos relacionados con el contraste de hipótesis.

Figura 3.5.4. Representación de la región de rechazo y p -valor del contraste de hipótesis de una distribución normal para $Z \geq 0$



3.5.2. Síntesis del análisis de los campos de problemas

En el análisis sobre los problemas de inferencia propuestos en las pruebas PAU, se ha observado que las prácticas matemáticas requeridas en su resolución implican el dominio de una gran variedad de conceptos, propiedades, lenguaje y procedimientos.

Tabla 3.5.6.*Configuración Epistémica Asociada al Campo de Problema C*

Tipos	Objeto matemático
Conceptos	Población y muestra, elementos Parámetro y estadístico Muestreo aleatorio simple Distribución muestral Probabilidad, asignación frecuencial Modelo teórico de distribución Distribución normal o binomial Distribución normal tipificada Teorema central del límite Hipótesis estadísticas, hipótesis nula y alternativa Contraste de hipótesis; contrastes unilaterales y bilaterales Regiones críticas y de aceptación, unilaterales y bilaterales Puntos críticos Nivel de significación p -valor Error tipo I y tipo II Probabilidad condicional
Propiedades	La muestra es una parte de la población Representatividad muestral Variabilidad muestral El parámetro es fijo y constante El estadístico es una variable aleatoria Linealidad de la media de la suma de variables aleatorias Linealidad de la varianza de la suma de v. aleatorias independientes Relación entre estadístico de la población y la distribución muestral Relación de la variabilidad muestral y tamaño de la muestra Condiciones de aplicación del teorema central del límite Aproximación normal a la distribución binomial Las hipótesis nula y alternativa son complementarias Las hipótesis se plantean respecto a los parámetros La hipótesis alternativa es la que interesa al investigador La hipótesis nula es la de no efecto Las regiones son complementarias No se puede calcular la probabilidad de la hipótesis La probabilidad del error tipo I es el nivel de significación La probabilidad del error tipo II es variable
Procedimientos	Tipificación Cálculo de probabilidades en la distribución muestral Lectura de las tablas de la $N(0,1)$ Manejo de desigualdades Aplicación de propiedades de la probabilidad Corrección de continuidad, para la distribución de la proporción Determinación de puntos críticos y regiones Rechazar o no la hipótesis Uso de la calculadora
Lenguaje Argumento	Verbal y simbólico, gráfico Deductivo Inductivo empírico

En la Tabla 3.5.7 se resumen los objetos matemáticos comunes cuya comprensión es evaluada en los diferentes campos de problemas considerados. Respecto a los conceptos, se comparten los de población, muestra y distribución muestral. Ello fuerza al estudiante a diferenciar los tres planos de distribución señalados por Schuyten (1991): 1) La distribución de los datos en la muestra; 2) La distribución de la variable de interés en la población; 3) La distribución muestral del estadístico en todas las posibles muestras obtenidas de la población.

Esta diferencia (y de los correspondientes resúmenes como la media o la varianza) es compleja y puede llevar al estudiante a cometer errores a la hora de diferenciar, por ejemplo, la varianza de la población y la de la distribución muestral de las medias al resolver los problemas. Globalmente y desde el punto de vista del número de conceptos implicados en su resolución, el problema más complejo es el tipo C, contraste de hipótesis, donde se comparten la mayoría de los conceptos con los otros campos y se añaden muchos más. Seguiría en complejidad los relacionados con el intervalo de confianza.

Tabla 3.5.7.

Conceptos Matemáticos Evaluados en los Diferentes Campos de Problemas

Conceptos	A1	A2	B1	B2	C
Parámetro y estadístico	X	X	X	X	X
Muestreo aleatorio simple	X	X	X	X	X
Muestreo estratificado, afijación proporcional	X	X			
Distribución muestral	X	X	X	X	X
Probabilidad, asignación clásica; principio de indiferencia	X	X			
Probabilidad, asignación frecuencial			X	X	X
Modelo teórico de distribución	X	X	X	X	X
Distribución uniforme discreta	X				
Distribución normal o binomial		X	X	X	X
Distribución normal tipificada		X	X	X	X
Teorema Central del Límite		X	X	X	X
Intervalo de confianza			X	X	
Coficiente de confianza o nivel de confianza			X	X	
Error de estimación				X	
Error máximo admisible				X	
Hipótesis estadísticas, hipótesis nula y alternativa					X
Contraste de hipótesis; contrastes unilaterales y bilaterales					X
Regiones crítica y de aceptación, unilaterales y bilaterales					X
Puntos críticos					X
Nivel de significación					X
<i>p</i> -valor					X
Error tipo I y tipo II					X
Probabilidad condicional					X

En la Tabla 3.5.8 se resumen las propiedades asociadas a cada campo de problema, donde las cinco primeras propiedades, así como la relación entre los estadísticos de la muestra, la población y la distribución muestral son compartidas. Asimismo, se observa otra serie de propiedades requeridas para todos los campos de problemas, excepto A1 y posteriormente se añaden propiedades específicas de los campos B1, B2 y C. Esta mayor densidad semiótica de conceptos y propiedades asociados a la resolución de problemas relacionados con el intervalo de confianza y contraste de hipótesis implica una mayor posibilidad de que el estudiante encuentre conflictos semióticos en la resolución, es decir, realice una interpretación inadecuada de algunos de los conceptos necesarios o asigne una propiedad que no le corresponde a los mismos.

Tabla 3.5.8.

Propiedades Evaluados en los Diferentes Campos de Problemas

Propiedades	A1	A2	B1	B2	C
La muestra es una parte de la población	X	X	X	X	X
Representatividad muestral	X	X	X	X	X
Variabilidad muestral	X	X	X	X	X
El parámetro es fijo y constante	X	X	X	X	X
El estadístico es una variable aleatoria	X	X	X	X	X
Linealidad de la media de la suma de variables aleatorias		X	X	X	X
Linealidad de la varianza de la suma de variables aleatorias independientes		X	X	X	X
Relación entre los estadísticos de la población y de la distribución muestral	X	X	X	X	X
Relación de la variabilidad muestral con el tamaño de la muestra		X	X	X	X
Condiciones de aplicación del teorema central del límite		X	X	X	X
Aproximación normal a la distribución binomial		X	X	X	X
Los extremos del intervalo de confianza son variables			X	X	
No se sabe si el parámetro está comprendido en el intervalo			X	X	
El coeficiente de confianza da el % de intervalos que cubren el parámetro			X	X	
Relación entre amplitud de intervalo, confianza y tamaño muestral			X	X	
Relación entre nivel de significación y coeficiente de confianza			X	X	
Probabilidad condicional			X		X
Las hipótesis se plantean respecto a los parámetros					X
La hipótesis alternativa es la que interesa al investigador					X
La hipótesis nula es la de no efecto					X
Las regiones son complementarias					X
No se puede calcular la probabilidad de la hipótesis					X
La probabilidad del error tipo I es el nivel de significación					X
La probabilidad del error tipo II es variable					X

Igualmente se requiere en todos los problemas el concepto de muestreo aleatorio simple, mientras el estratificado aparece solo en los campos A1 y A2. Aunque es necesaria la idea de probabilidad en todos los casos, en general se aplica la visión frecuencial de la misma, excepto en los campos A1 y A2 que requiere de la probabilidad clásica. Los modelos clásicos de distribución binomial y normal van asociados a los problemas A2, B1, B2 y C que también tienen asociados la distribución normal tipificada y el teorema central del límite y la distribución uniforme al A1. El resto de conceptos son prototípicos de los campos B1, B2 y C y se añaden a todos los anteriores, lo que implica una mayor complejidad semiótica de estos tres últimos campos de problemas en relación con los problemas A1 y A2.

Es menor la diferencia respecto al resto de objetos matemáticos. Si que hay una clara diferencia en los procedimientos requeridos para el campo de problemas A1 y el resto, pero los otros cuatro campos comparten muchos de los procedimientos (véase Tabla 3.5.8). Esto quizás explica que los profesores concentren el poco tiempo disponible para preparar a los estudiantes para las pruebas en la enseñanza de procedimientos, pues en principio, es posible aplicar estos procedimientos en forma mecánica y la comprensión se observará sólo cuando se pida interpretar los resultados.

Todos los campos de problemas propuestos se apoyan fuertemente en la comprensión de la probabilidad condicional, que interviene tanto en la definición de las distribuciones muestrales (condicionadas al valor del parámetro), intervalos de confianza (condicionados con el valor del estadístico) como en el contraste de hipótesis, donde las regiones de aceptación y rechazo se calculan bajo la condición de ser cierta la hipótesis nula (Castro-Sotos et al., 2007). Puesto que el concepto de probabilidad condicional es difícil y se han descrito números sesgos sobre el mismo (Díaz-Batanero et al., 2012; Díaz y de la Fuente, 2005; Falk, 1986) los profesores deberán prestar especial atención a su uso en inferencia estadística, para asegurar un aprendizaje adecuado por parte de los estudiantes y garantizar una correcta solución de los problemas.

La Tabla 3.5.9 muestra otros objetos matemáticos implicados en los distintos campos de problemas. Se observa que todos los campos comparten el lenguaje verbal, simbólico y a veces gráfico que se debe dominar, así como el razonamiento deductivo al

que se añade el razonamiento inductivo empírico en algunos de los problemas (véase Tabla 3.5.9).

Tabla 3.5.9.

Otros Objetos Matemáticos Evaluados en los Diferentes Campos de Problemas

		A1	A2	B1	B2	C
Procedimientos	Enumerar las muestras de una población finita	X				
	Formar muestras de una población finita	X				
	Cálculo de los estadísticos en las muestras	X				
	Calcular los estadísticos de la distribución muestral	X				
	Cálculo del tamaño de la muestra	X				
	Tipificación		X	X	X	X
	Cálculo de probabilidades en la distribución muestral		X	X	X	X
	Lectura de las tablas de la N (0,1)		X	X	X	X
	Manejo de desigualdades		X	X	X	X
	Aplicación de propiedades de la probabilidad		X	X	X	X
	Corrección de continuidad, para el caso de la distribución de la proporción		X	X	X	X
	Cálculo de extremos del intervalo de confianza			X	X	
	Cálculo del tamaño mínimo de muestra para una precisión dada				X	
	Determinación de puntos críticos y regiones					X
	Rechazar o no la hipótesis					X
	Uso de la calculadora			X	X	X
	Lenguaje	Verbal y simbólico, gráfico	X	X	X	X
Argumento	Deductivo	X	X	X	X	
	Inductivo empírico			X	X	X

De la información recogida en la Tabla 3.5.7, se observa que, puesto que la construcción de intervalos de confianza y el contraste de hipótesis se apoyan sobre los conceptos tratados en A1 y A2 (comprensión del muestreo y distribución muestral), en general, los tres campos, B1, B2 y C, incluyen más contenidos evaluados que los A1 o A2. Esto implica una mayor complejidad de este tipo de problemas respecto a los referidos exclusivamente a muestreo y distribución muestral.

En cuanto a los procedimientos, destacamos la importancia que tiene una correcta lectura de la tabla de la distribución normal tipificada, ya que es necesario su uso en todos los campos a excepción de los problemas A1. Señalar también la necesidad de realizar correctamente los cálculos estadísticos con el fin de obtener las medidas de centralización, medidas de dispersión y proporción muestral para, posteriormente, describir la distribución tanto de la media muestral como de la proporción muestral.

Por último, otro punto importante a tener en cuenta es el manejo de desigualdades en el proceso de tipificación.

3.5.3. Estudio comparativo entre el significado pretendido y evaluado

Como se indicó en la introducción, es importante asegurar la correspondencia entre el significado institucional que el currículo propone para los objetos matemáticos (en este caso la inferencia estadística) y el significado evaluado, en este caso en las PAU. Para estudiar esta correspondencia, en la Tabla 3.5.10, se realiza una descomposición de los contenidos sobre inferencia estadística propuestos para Matemáticas Aplicadas a las Ciencias Sociales II en el Decreto de Enseñanza Mínimas (MEC, 2007, p. 45476) que estuvo vigente en la mayoría de las pruebas que han sido analizadas. Se observa que los contenidos fijados en el significado previsto de la inferencia estadística en las directrices curriculares citadas se evalúan en algunos de los campos de problemas planteados, aunque no con la misma intensidad. Así los contenidos relacionados con el teorema central del límite y la aproximación a la distribución normal serán empleados siempre que se trabaje con la distribución binomial para obtener distribuciones muestrales, calcular o interpretar intervalos de confianza y plantear contrastes de hipótesis referidos a la proporción en una población binomial.

Por otro lado, los problemas relacionados con la elección de la muestra sólo se evalúan en los campos de problemas relacionados con ellas, es decir, sobre tipos de muestreo. El conocimiento sobre intervalo de confianza o contraste de hipótesis, sus propiedades y los procedimientos para llevarlos a cabo se evalúan en los campos de problemas B o C respectivamente y el conocimiento de la distribución muestral en todos los campos propuestos.

La Tabla 3.5.11 recoge el análisis respecto a los contenidos de inferencia estadística, a través de los estándares de aprendizaje evaluables para la asignatura de Matemáticas Aplicadas a las Ciencias Sociales II en el último decreto curricular (MECD, 2015b, p. 389). Se comparan estos contenidos con los evaluados en cada uno de los campos de problemas analizados. Se observa que los intervalos de confianza tienen mayor peso en estos estándares y el contraste de hipótesis solo se recoge en uno de ellos. También se aprecia que tanto el cálculo de estimadores puntuales (media,

varianza, desviación típica y proporción poblacionales), como el cálculo de probabilidades asociadas a la distribución de la media muestral y de la proporción muestral se recogen en todos los campos de problemas analizados. Nótese que el resto de estándares de aprendizaje son específicos del intervalo de confianza o el contraste de hipótesis. Por último, señalar que, aunque la representatividad de la muestra es un elemento de gran importancia dentro de la inferencia estadística, ninguno de los campos de problemas recoge el estándar de la valoración de la representatividad de la muestra.

Tabla 3.5.10.

Descomposición del Criterio de Evaluación sobre Inferencia para la Asignatura Matemáticas Aplicadas a las Ciencias Sociales II (MEC, 2007, p. 45477)

Criterios de Evaluación	Campos de problemas				
	A1	A2	B1	B2	C
Implicaciones prácticas de los teoremas: Central del Límite, Aproximación de la Binomial a la Normal y Ley de los Grandes Números.		X	X	X	X
Problemas relacionados con la elección de las muestras. Condiciones de representatividad.	X	X			
Parámetros de una población	X	X	X	X	X
Distribuciones de probabilidad de las medias y proporciones muestrales.		X	X	X	X
Intervalo de confianza para el parámetro p de una distribución binomial y para la media de una distribución normal de desviación típica conocida.			X	X	
Contraste de hipótesis para la proporción de una distribución binomial y para la media o diferencias de medias de distribuciones normales con desviación típica conocida					X

3.6. MODELO PROBABILÍSTICO DE LA POBLACIÓN DE PARTIDA

Recordemos que la estadística inductiva o inferencia estadística se centra en la búsqueda de herramientas las cuales nos permiten obtener conclusiones sobre una población a partir del estudio estadístico realizado a una muestra (Batanero y Díaz, 2006). Tanto a la hora de seleccionar una muestra, mediante las distintas técnicas de muestreo, como en el proceso de estimación de la media y la proporción poblacional, se parte del estudio de una variable aleatoria en la población de partida, que se distribuye según un modelo teórico determinado. Cada uno de estos modelos queda determinado por unos parámetros sobre los cuáles se centra la inferencia paramétrica. Por tanto, a

la hora de comenzar el estudio es necesario tener bien claro el concepto de población, la variable aleatoria estudiada y los parámetros que la caracterizan, es decir, será necesario tenerla perfectamente definida (Rossman, 2008).

Tabla 3.5.11.

Estándares sobre Inferencia para la Asignatura Matemáticas Aplicadas a las Ciencias Sociales II (MECD, 2015b, pp. 389)

Estándares sobre inferencia	Campos de problema				
	A1	A2	B1	B2	C
1.4. Resuelve una situación relacionada con la toma de decisiones en condiciones de incertidumbre en función de la probabilidad de las distintas opciones.					X
2.1. Valora la representatividad de una muestra a partir de su proceso de selección.					
2.2. Calcula estimadores puntuales para la media, varianza, desviación típica y proporción poblacionales, y lo aplica a problemas reales.	X	X	X	X	X
2.3. Calcula probabilidades asociadas a la distribución de la media muestral y de la proporción muestral, aproximándolas por la distribución normal de parámetros adecuados a cada situación, y lo aplica a problemas de situaciones reales.	X	X	X	X	X
2.4. Construye, en contextos reales, un intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida.			X		
2.5. Construye, en contextos reales, un intervalo de confianza para la media poblacional y para la proporción en el caso de muestras grandes.			X		
2.6. 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.				X	
3.1. Utiliza las herramientas necesarias para estimar parámetros desconocidos de una población y presentar las inferencias obtenidas mediante un vocabulario y representaciones adecuadas.			X	X	X

Dado que los problemas propuestos en las PAU integran diversos modelos probabilísticos, se ha analizado la distribución o modelo probabilístico de la población de partida para analizar si dichos modelos corresponden a las distribuciones estudiadas y propuestas en las directrices curriculares.

En los problemas de los campos A1 y A2, el interés es el muestreo y a veces no se explicita en el enunciado cuál es la distribución subyacente. Por tanto, el estudiante ha de recordar que en la realización de un muestreo aleatorio *a priori* se supone que la

probabilidad de ser elegidos es la misma para todos y consecuentemente la población, de donde ha sido extraída dicha muestra, tendrá asociado el modelo probabilístico uniforme discreto. Ejemplo de ello son los dos problemas propuestos en la prueba de reserva 2B del curso 2013-2014.

Problema 2B (2014)

- 1) En un centro docente la tercera parte de los alumnos estudia el idioma A, la mitad el idioma B y el resto el idioma C (cada alumno estudia sólo uno de estos idiomas).
 - a) Se desea seleccionar una muestra de 60 alumnos, mediante muestreo aleatorio estratificado con afijación proporcional al número de los alumnos de cada idioma. ¿Cómo debería estar conformada la muestra?
 - b) En otra muestra seleccionada por el procedimiento anterior, el número de alumnos tomados del idioma A es 14. Determine cuántos se han elegido de los otros dos idiomas.
- 2) Una población tiene 5 elementos. Mediante muestreo aleatorio simple se seleccionan muestras de tamaño 3, siendo la desviación típica de sus medias 2 y la media de las medias muestrales 7. ¿Cuánto valen la media y la varianza de la población?

En el primero no se especifica el tamaño de la población, pero se ha de suponer que todos los alumnos tienen la misma probabilidad de entrar en la muestra; por tanto, la distribución es uniforme. En el 2) se trata de una pequeña población con tan sólo cinco elementos a la que también ha de suponerse una distribución uniforme discreta. Dicha distribución quedará especificada por la probabilidad asociada a cada elemento de la población.

En el caso de realizar alguna inferencia sobre el valor de la media poblacional se supone que la población sigue una distribución normal con desviación típica conocida, que se da como dato del problema. En esta situación, si la población de partida es normal con varianza conocida, entonces la distribución muestral de las medias es también normal, debido al hecho de que la distribución de la suma de variables normales independientes es normal.

Los parámetros de la distribución muestral se calcularán teniendo en cuenta la linealidad de la media de cualquier variable aleatoria (la media de la suma es igual a la suma de las medias) y también que en el caso de que las variables que se suman sean independientes, la varianza de la suma es igual a la suma de las varianzas. Estas propiedades no siempre se explicitan en la enseñanza, sino que se dan como hechos asumidos, para los estudiantes, que los aprenden como recetas y sin significado alguno.

En el ejemplo problema 5A de 2009 se facilita al estudiante la información referente a la distribución de la población en estudio que es necesario en la construcción del intervalo de confianza.

Problema 5A (2009)

Una variable aleatoria X se distribuye de forma Normal, con media μ y desviación típica $\sigma=0,9$.

a) Una muestra aleatoria de tamaño 9 ha proporcionado los siguientes valores de X : 7,0; 6,4; 8,0; 7,1; 7,3; 7,4; 5,6; 8,8; 7,2. Obtenga un intervalo de confianza para la media μ , con un nivel de confianza del 97%.

b) Con otra muestra, se ha obtenido que un intervalo de confianza para μ , al 95%, es el siguiente (6,906, 7,494). ¿Cuál es el tamaño de la muestra utilizada?

En el caso de los problemas relacionados con la estimación de la proporción poblacional, se supone que la distribución de la población de partida es una distribución binomial. Este tipo de problema suele revestir mayor dificultad, puesto que, por un lado, no se da el valor de la varianza o desviación típica y es el estudiante quien debe recordar la fórmula para determinarla, y, por otro lado, puesto que las muestras suelen ser grandes, hay que utilizar la aproximación de la distribución binomial a la distribución normal. Ejemplo de ello es el problema 6A (2011) en el que el estudiante debe identificar la distribución de la población, ya que no es facilitada como dato. También debe recordar las condiciones de aproximación de la distribución binomial por la normal y aplicar el teorema central del límite, puesto que la muestra tiene un tamaño grande, para aceptar que la distribución muestral se distribuye normalmente. Por tanto, el campo de problemas correspondiente a la distribución binomial supone mayor dificultad al tener que recordar el estudiante más conceptos y propiedades al resolverlo.

Problema 6A (2011)

El director de una televisión afirma que un nuevo programa que va a emitirse será visto, al menos, por un 30% de personas. Una vez emitido se realizó una encuesta a 500 personas, elegidas al azar, y ésta reveló que 130 de ellas habían visto ese programa.

a) Formule la hipótesis nula y la alternativa del contraste de hipótesis que permite determinar si los datos de la encuesta realizada son compatibles con la afirmación del director.

b) Halle la región crítica de ese contraste para un nivel de significación del 5.5%.

c) Según el dato obtenido en el apartado anterior ¿qué conclusión se obtiene sobre la afirmación realizada por el director de esa televisión?

Realizando el estudio a partir de los campos de problemas, a partir de la información de la Tabla 3.6.1, se observa que a excepción del campo A1, donde la distribución de la población es uniforme discreta, en el resto de los campos (distribución

muestral en poblaciones infinitas, cálculo de probabilidades, construcción e interpretación de intervalos de confianza y contraste de hipótesis) los modelos probabilísticos de la población son binomial o normal. De hecho, además, en la práctica, siempre la distribución binomial se aproxima por la normal, lo que indica que las PAU dan prioridad a este modelo de distribución.

Tabla 3.6.1.

Modelos Probabilísticos Incluidos en los Campos de Problema

	Campos de problema				
	A1	A2	B1	B2	C
Distribución Uniforme Discreta	X				
Distribución Binomial		X	X	X	X
Distribución Normal		X	X	X	X

Por último, se debe señalar que, aunque todos los campos, a excepción del A1, trabajan tanto con la distribución normal como con la binomial, en general, y tal como se ha puesto de manifiesto en el análisis sobre los procedimientos recogidos en la Tabla 3.5.7, en la mayor parte de los campos de problemas se termina recurriendo al uso de la tabla de la distribución normal, la tipificación y el cálculo de probabilidades basadas en dicha distribución. Por lo que se puede se pone de manifiesto la importancia que tiene el uso distribución normal dentro de la inferencia estadística y en las pruebas de acceso a la universidad.

3.7. PARÁMETRO POBLACIONAL OBJETO DE ESTUDIO

En el campo de la inferencia estadística los métodos básicos son la estimación y los contrastes de hipótesis. Obtenida una muestra representativa de la variable aleatoria en ciertas ocasiones el problema a resolver puede estar centrado en identificar las características del modelo en relación a su forma (contrastar el modelo probabilístico de la población, entre otros) o realizar estimaciones sobre los parámetros de los que depende el modelo probabilístico. Dado que tanto las normativas curriculares como las PAU ponen el foco de atención en la inferencia paramétrica, en este apartado se han analizado los parámetros del modelo probabilístico en cuestión. Concretamente se ha examinado aquellos parámetros que deben ser estimados mediante estimación puntual

(a partir de los estadísticos muestrales, se asigna un valor al parámetro poblacional), estimación por intervalo de confianza (intervalo dentro del cual se espera este contenido el verdadero valor del parámetro desconocido con cierto grado de confianza) y aquellos sobre los que se plantea una hipótesis a contrastar. No todos los problemas solicitan la estimación del mismo parámetro o bien la realización de un contraste sobre el mismo y, aunque no siempre, generalmente el tipo de parámetro que se pide se relaciona con la distribución en la población.

En general los parámetros a estimar en cada uno de los campos de problemas (véase Tabla 3.7.1) son los referentes a la media poblacional y la proporción poblacional. En cada caso se trabaja con el correspondiente estadístico muestral (media o proporción) y el estudiante ha de diferenciarlo respecto al parámetro (Harradine et al., 2011). Con respecto a la estimación de la varianza poblacional, solamente es posible encontrarla en los problemas asociados a la categoría A1 (composición de las muestras y cálculos estadísticos de las distribuciones muestrales en poblaciones finitas) ya que, en todos los problemas relacionados con la estimación de media poblacional siempre se parte de que la varianza es conocida en la población. Se trabaja con la varianza muestral, aunque en este caso hay un sesgo de significado, pues el estimador insesgado de la varianza en la población es la cuasivarianza muestral, donde, en lugar de dividir por el tamaño de la muestra, se divide por el tamaño de la muestra menos uno.

Tabla 3.7.1.

Clasificación de los parámetros a estimar según los campos de problema

	Campos de problema				
	A1	A2	B1	B2	C
Proporción	X	X	X	X	X
Media	X	X	X	X	X
Varianza/Desviación típica	X				

A continuación, se muestra un ejemplo para cada uno de los tres tipos de parámetros diferentes: media poblacional (problema 1A curso 2011-2012), varianza poblacional (problema 3A del curso 2003-2004) y proporción poblacional (problema 3B curso 2005-2006).

Problema 1A (2012)

La variable “tiempo de reacción de un conductor ante un obstáculo imprevisto” sigue una distribución Normal con desviación típica 0,005 segundos. Al medir dicho tiempo en 50 conductores se ha obtenido un tiempo medio de 0,85 segundos.

- a) Halle el intervalo de confianza para el tiempo medio de reacción, con un nivel de confianza del 99%.
- b) ¿De qué tamaño mínimo ha de tomarse una muestra para que el error de estimación no supere 0,01 segundos, con un nivel de confianza del 95%??

Problema 3A (2004)

Dada la población de elementos {3, 4, 5, 8}, se pretende seleccionar una muestra de tamaño 2, mediante muestreo aleatorio con reemplazamiento.

- a) Escriba todas las muestras posibles.
- b) Calcule la varianza de la población.
- c) Calcule la varianza de las medias muestrales.

Problema 3B (2006)

Se ha lanzado un dado 400 veces y se ha obtenido 80 veces el valor cinco. Estime, mediante un intervalo de confianza al 95 %, el valor de la probabilidad de obtener un cinco.

3.8. CONTEXTO DEL PROBLEMA

El contexto en el que se enmarca un problema es esencial para mostrar la unión existente entre los conceptos y las situaciones reales, permitiendo darle sentido al aprendizaje y motivar el interés por las matemáticas, haciéndoles ver su utilidad en diferentes contextos. Dada la naturaleza abstracta que tiene las matemáticas, el uso de un contexto cercano al alumno es una herramienta que adquiere gran importancia para garantizar un proceso de enseñanza-aprendizaje más eficiente.

El análisis de este elemento adquiere cierta importancia ya que, tal y como señala Rico (2006b), el uso de las matemáticas en situaciones y contextos variados es un componente importante de la alfabetización matemática. Las situaciones permiten conectar un problema con los fenómenos de los que surgen los objetos matemáticos considerados.

A continuación, se presenta cada uno de los contextos junto a un ejemplo extraído de las pruebas PAU. En el análisis se han contemplado los contextos recogidos en las pruebas de evaluación PISA (OCDE, 2015), orientadas a evaluar la competencia matemática, es decir, el contexto personal, profesional, social y científico. Aquellos

problemas que no pueden ser incluidos en alguno de los contextos descritos se consideran problemas descontextualizados o problemas sin contexto.

Situación personal. Se asocian los problemas relacionados con las actividades del día a día del propio individuo, su familia y su grupo de iguales. Dicho contexto será muy cercano para el estudiante, pudiendo utilizar su propio conocimiento para resolver el problema. Como ejemplo, se muestra el problema 3B propuesto en 2004, pues el contexto es el número de horas que dedica el estudiante a la práctica de deporte, que es una variable personal para el estudiante. Otros ejemplos de problemas en contexto personal son los relacionados con las compras, familia, juegos, salud o transporte personal, deportes, o viajes.

Problema 3B (2004)

El número de horas semanales que los estudiantes de Bachillerato de una ciudad dedican al deporte se distribuye según una ley Normal de media 8 y varianza 7,29.

- a) Para muestras de tamaño 36, indique cuál es la distribución de las medias muestrales.
- b) ¿Cuál es la probabilidad de que la media de una muestra de tamaño 36 esté comprendida entre 7,82 y 8,36 horas?

Situación profesional: Son problemas que se centran en el mundo laboral que el estudiante encontrará en el futuro o que conoce por sus padres o maestros. Se trata de problemas sobre medida, control o coste de un proceso de producción o una construcción, beneficios, salarios o ventas de una empresa, sobre diseño en carpintería, arquitectura o jardinería; coste o salario de mano de obra, etc. El interés de que el estudiante trabaje con este tipo de problemas es, por un lado, mostrarle la utilidad de la matemática fuera del aula, y por otro concienciarle de que lo que aprende le puede ser de utilidad posterior. En nuestro caso se ha considerado como ejemplo el problema 1A propuesto en 2007 siendo el salario de los trabajadores de una ciudad la variable aleatoria bajo estudio.

Problema 1A (2007)

El salario de los trabajadores de una ciudad sigue una distribución Normal con desviación típica 15 euros. Se quiere calcular un intervalo de confianza para el salario medio con un nivel de confianza del 98%. Determine cuál es el tamaño mínimo de la muestra que se necesitaría recoger para que el intervalo de confianza tenga una amplitud, como máximo, de 6 euros.

Situación Social: Se contemplan los problemas que el estudiante podría encontrar en su comunidad más amplia que la familiar (comunidad de vecinos, ayuntamiento o ciudad, su país, etc.). En este contexto pueden ser incluidos también, por ejemplo, los sistemas electorales, las estadísticas nacionales, el transporte urbano o interurbano, la demografía y la economía. Dentro de este tipo se han incluido también los juegos de azar como la lotería, etc., que aparecen en los medios de comunicación. El problema 1B de 2012 se ha incluido dentro de esta categoría por tratar el tema del uso del carril bici en una ciudad.

Problema 1B (2012)

Un informe de un Ayuntamiento afirma que al menos el 26% de los usuarios del carril bici habrían utilizado el coche particular para sus desplazamientos de no haber existido dicho carril. Sin embargo, un periódico local anuncia la falsedad del dato, informando que una encuesta propia indica que solo 240 de los 1000 usuarios encuestados afirman que habrían utilizado el coche particular.

- a) Establezca un contraste, con hipótesis nula, para verificar la afirmación del Ayuntamiento e indique la región crítica de dicho contraste para un nivel de significación del 5%.
- b) Con este nivel de significación, ¿podría aceptarse el informe del Ayuntamiento?

Situación Científica: Los problemas clasificados en la categoría científica están relacionados con la aplicación de las matemáticas en ciencia y tecnología y la investigación, en general. A modo de ejemplo se muestra el problema 6A propuesto en 2013, cuyo objetivo es realizar un estudio sobre la proporción de hembras entre los peces de una piscifactoría. Otros problemas relacionados con la ciencia tratan de la biología, botánica, meteorología, ecología, medicina, genética, psicología o física. Esta situación es más abstracta que el resto ya que implica la comprensión de un proceso tecnológico, una interpretación teórica o un problema matemático.

Problema 6A (2013)

Se quiere estimar la proporción de hembras entre los peces de una piscifactoría; para ello se ha tomado una muestra aleatoria de 500 peces, y en ella hay 175 hembras.

- a) Calcule un intervalo de confianza para la proporción de hembras en esta población de peces, con un nivel de confianza del 94%.
- b) A la vista del resultado del muestreo se quiere repetir la experiencia para conseguir un intervalo de confianza con el mismo nivel y un error máximo de 0,02, ¿cuál es el tamaño mínimo que debe tener la nueva muestra?

Sin contexto. Se han contemplado aquellos problemas que no han sido incluidos en ninguno de los contextos mencionados anteriormente. Concretamente, se han

incluido en esta categoría aquellos enunciados redactados sobre situaciones matemáticas abstractas y en los que no se incluye ninguna aplicación a la vida real. Ejemplo de ello es el problema 5A propuesto en 2012. Señalar que la inclusión de este tipo de ejercicios contradice todas las recomendaciones sobre la enseñanza de la estadística en que el contexto es fundamental para el aprendizaje.

Problema 5A (2012)

Una característica de una determinada población se distribuye según una variable aleatoria Normal X de media desconocida y desviación típica 0,9. Extraída al azar una muestra de tamaño 9 de esa población y observada X , dio como resultados: 10,5; 10; 8,5; 10,5; 11,5; 13,5; 9,5; 13; 12.

a) Halle un intervalo de confianza, al 99%, para la media de la variable X .

b) Determinar el tamaño mínimo que debe tener una muestra de esa población, para que el error máximo que se cometa en la determinación de un intervalo de confianza para la media de X sea, a lo sumo, 0,3, con un nivel de confianza del 90%.

3.9. ANÁLISIS ESTADÍSTICO DE LOS PROBLEMAS

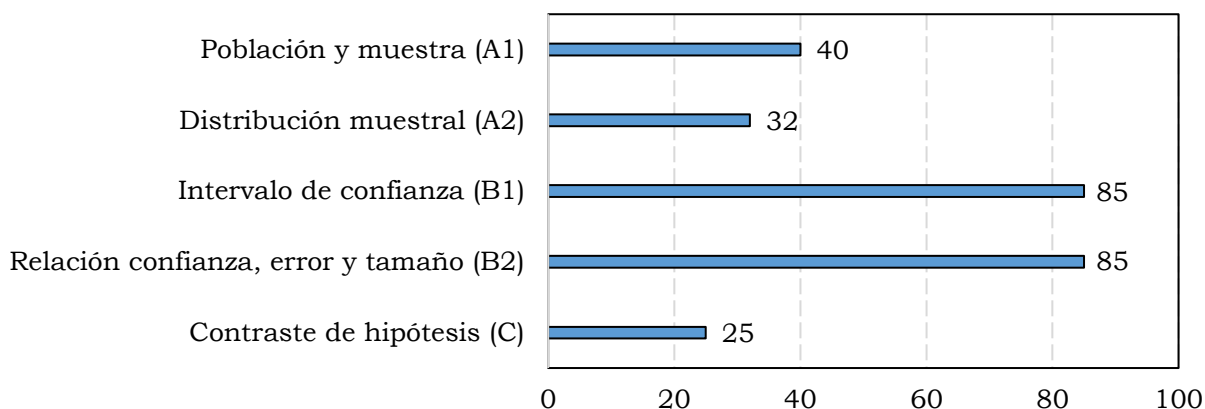
A partir del análisis cualitativo descrito en los anteriores apartados, donde se identificaron las categorías para cada una de las variables y para cada problema, se han codificado los datos para su posterior análisis estadístico. En lo que sigue, se presenta la distribución, globalmente y por año, de las categorías en la muestra de problemas analizados. Los resultados serán presentados siguiendo el mismo orden de análisis realizado para cada una de las variables en las Secciones 3.5, 3.6, 3.7 y 3.8.

3.9.1. Campos de problemas

Tal y como se ha mencionado anteriormente los campos de problemas que se han tratado son: A) muestreo (dividido en las categorías A1 y A2); B) intervalos de confianza (dividido en las categorías B1 y B2); C) contrastes de Hipótesis. En la Figura 3.9.1 se muestra la distribución global de los campos de problemas propuestos en las 144 pruebas analizados. Se ha comprobado que un mismo problema puede contener apartados correspondientes a distintos campos de problemas. Este hecho nos indica que prácticamente en todos los problemas se ha tratado de evaluar el conocimiento de más de un campo de problemas de entre los que se han definido.

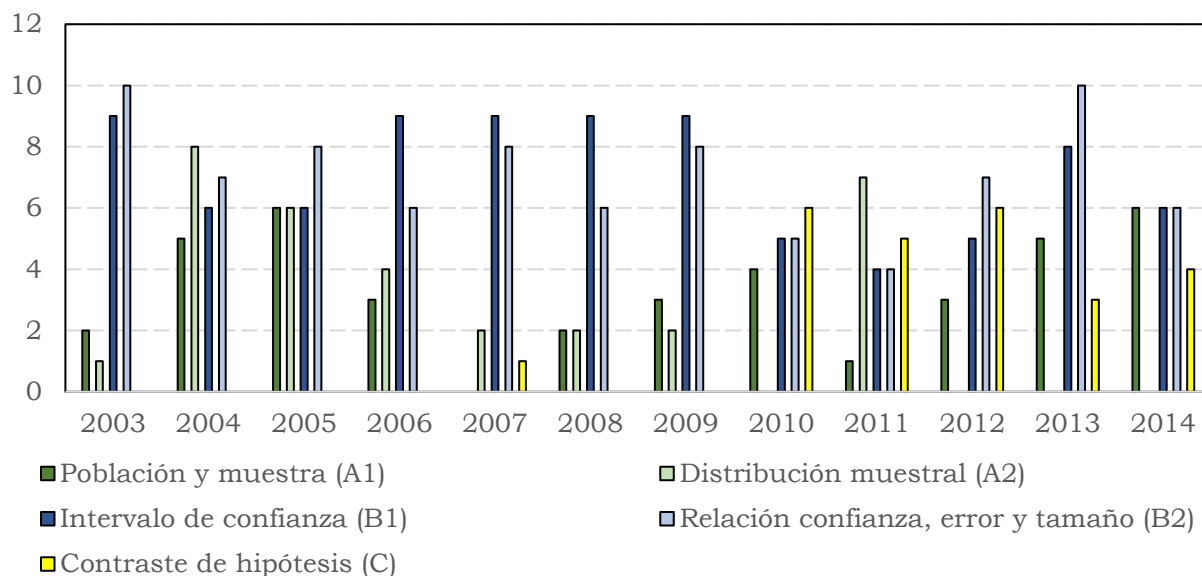
Se observa que los campos de problemas que aparecen con mayor frecuencia en las pruebas son los relacionados con la construcción e interpretación de los intervalos de confianza y aquellos que hacen referencia a la relación entre la confianza, el error de estimación y el tamaño muestral. Conjuntamente, los problemas relacionados con el intervalo de confianza (campos B1 y B2) suponen aproximadamente dos tercios (64%) del total de problemas planteados, por lo que se deduce la gran importancia dada a este tipo de problemas en las PAU. En este sentido, y teniendo en cuenta que en estos problemas también se evalúa la comprensión de la distribución muestral y sus estadísticos, hay un buen ajuste entre los problemas propuestos en las PAU y los estándares de evaluación previstos en el actual desarrollo curricular (MECD, 2015b). Sin embargo, el contraste de hipótesis, previsto en el desarrollo curricular anterior (MEC, 2007) quedaría infrarrepresentado en las PAU.

Figura 3.9.1. *Frecuencia de apartados según campo de problema*



Por otro lado, si se analizan los distintos tipos de problemas propuestos en cada uno de los años (véase Figura 3.9.2) se observa que no es hasta el curso académico 2007-2008 donde se incorporan los problemas de contrastes de hipótesis. A partir de su inclusión en las PAU, casi la mitad de los problemas propuestos a partir de entonces corresponden a este campo de problema. La aparición de los contrastes de hipótesis supuso en el periodo analizado una disminución de los problemas relacionados con el campo de problema B (pasa del 68% hasta 2009 al 58% desde 2010), indicando un cambio de tendencia en los contenidos de inferencia evaluados en las pruebas de acceso.

Figura 3.9.2. Distribución de la variable campos de problemas en el periodo 2003-2014



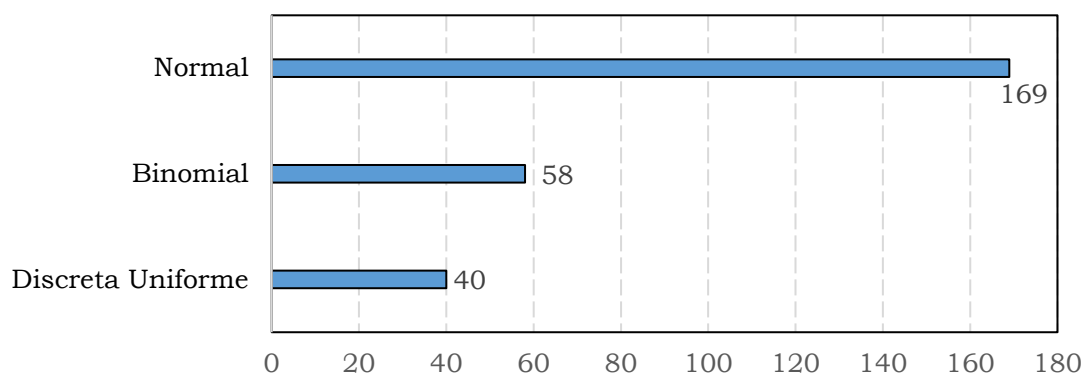
Finalmente se resalta el hecho de que los problemas relacionados exclusivamente con muestreo (categorías A1 o/y A2) tienen una distribución variable. A lo largo del periodo seleccionado se observa que todos los años incluyen problemas relacionados con la categoría A1, a excepción del 2007. Sin embargo, la categoría A2, aunque en algunos años ha tenido una presencialidad destacada, por ejemplo, en 2004 y 2005, la tendencia en los últimos cinco años, a excepción de 2011, ha sido eliminarlos de las pruebas.

3.9.2. Modelo probabilístico de la población de partida

El análisis de los modelos probabilísticos que sigue la población en estudio muestra la importancia que tiene dentro de la inferencia estadística la distribución normal. La Figura 3.9.3 recoge la frecuencia de los apartados según las tres distribuciones de probabilidad identificadas en las PAU. Se observa que la distribución normal se incluye aproximadamente en el 63% de los problemas analizados. Por otro lado, aunque generalmente a posteriori se suele realizar la aproximación de la distribución binomial a la normal, los problemas relacionados con proporciones nos han permitido concluir que el 22% de ellos parten de una población binomial. En menor medida se han

detectado problemas en los que la distribución uniforme de tipo discreta es el modelo de la población de partida pues solamente se ha encontrado en el 15% de los casos.

Figura 3.9.3. Frecuencia de apartados según modelo probabilístico de la población de partida

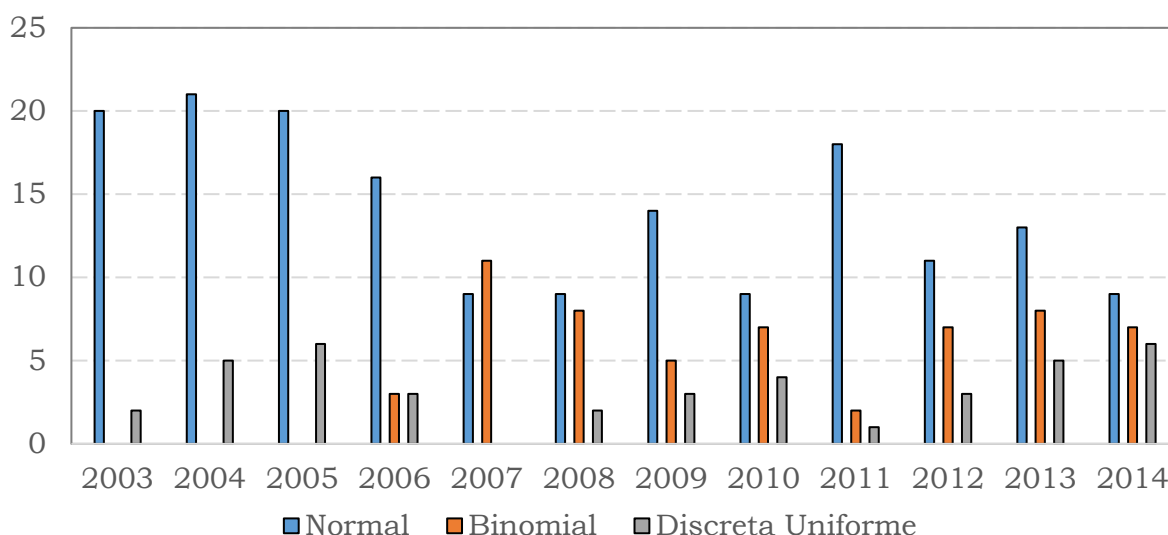


Analizando los problemas cada año (véase Figura 3.9.4), y comparándolo con los resultados obtenidos en el punto anterior, se comprueba que el hecho de no incluir problemas de la categoría A1 en 2007 implica la no inclusión de la distribución uniforme discreta. De la Figura 3.9.4, se puede concluir que no es hasta el 2006 donde se introducen ejercicios donde el parámetro en estudio sea la proporción muestral, ya que en años previos no se ha contabilizado ningún problema donde la población se distribuya según el modelo binomial. La introducción de este modelo probabilístico supone un gran cambio, en el periodo 2003-2005 el 90% de los problemas analizados partían de un población normal, sin embargo, en el resto de años este porcentaje disminuyó hasta el 56% considerando el modelo binomial en el 30% de los problemas propuestos en los últimos nueve años.

A diferencia del resto de campos de problemas, se ha detectado que la mitad de los problemas relacionados con contraste de hipótesis no están subdivididos en apartados (11 de los 25 problemas). Este hecho acentúa la dificultad de este tipo de problemas ya que el estudiante se juega toda la calificación (2,5 puntos) en único problema. Además, con los problemas de contraste de hipótesis aumenta la dificultad pues autores como Fidler y Cumming (2005) indican que la comprensión de los intervalos es más sencilla que la de los contrastes de hipótesis tanto para estudiantes

como para investigadores. Adicionalmente, contradice el nuevo currículo en el que se ha suprimido el contraste de hipótesis.

Figura 3.9.4. Distribución de la variable modelo probabilístico de la población de partida en el periodo 2003-2014



3.9.3. Parámetro poblacional objeto de estudio

La Figura 3.9.5 muestra los distintos parámetros a estimar junto a su frecuencia. En la codificación de esta variable se informa que la categoría *sin parámetro* hace referencia a aquellos problemas en los que se pide explícitamente la distribución de la media muestral mediante el uso del teorema central del límite y el posterior cálculo de una probabilidad, así como aquellos en los que se solicita la composición o enumeración de las muestras.

Se observa que la media poblacional es el parámetro a estimar que además de ser el parámetro con mayor peso, es el único que se ha contabilizado en todos los años del periodo seleccionado PAU (55%), resultado que va en la misma línea al obtenido en la Sección 3.9.2. En relación al parámetro varianza únicamente se ha encontrado en nueve problemas a lo largo de los tres primeros años seleccionados en el estudio (véase Figura 3.9.6) pasando casi desapercibida en las pruebas. Con la supresión de la varianza se incorpora la proporción poblacional manteniéndose hasta el 2014 implicando una disminución de los problemas sobre la media poblacional.

Figura 3.9.5. Frecuencia de la variable parámetro a estimar

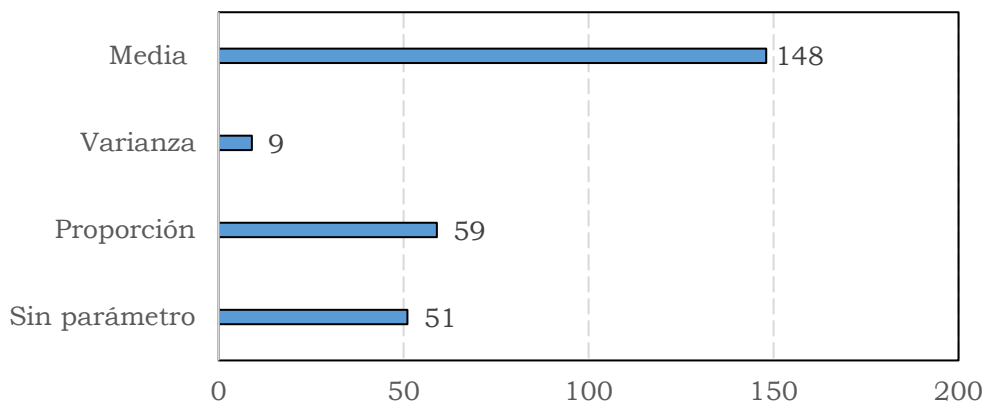
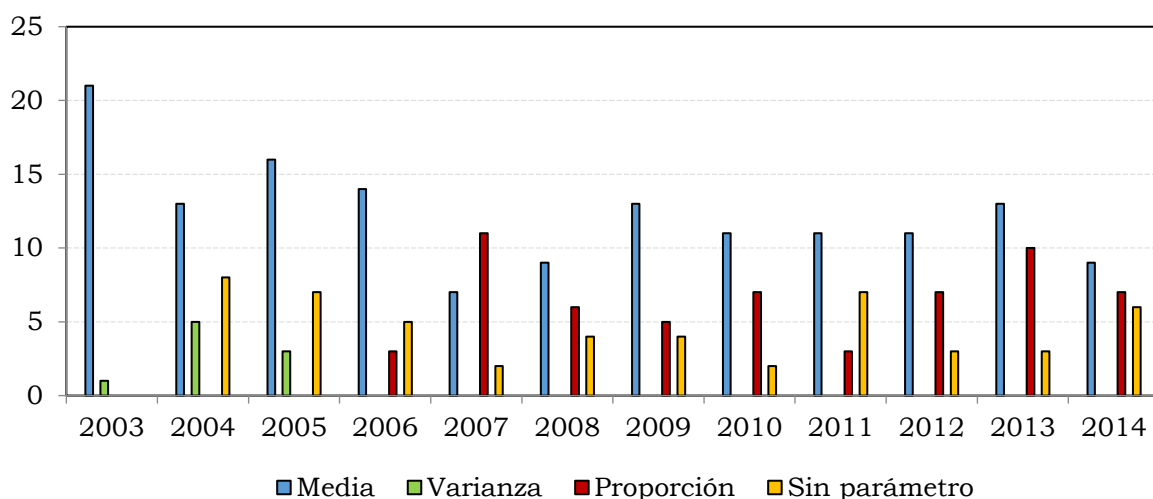


Figura 3.9.6. Distribución de la variable parámetro a estimar en el periodo 2003-2014

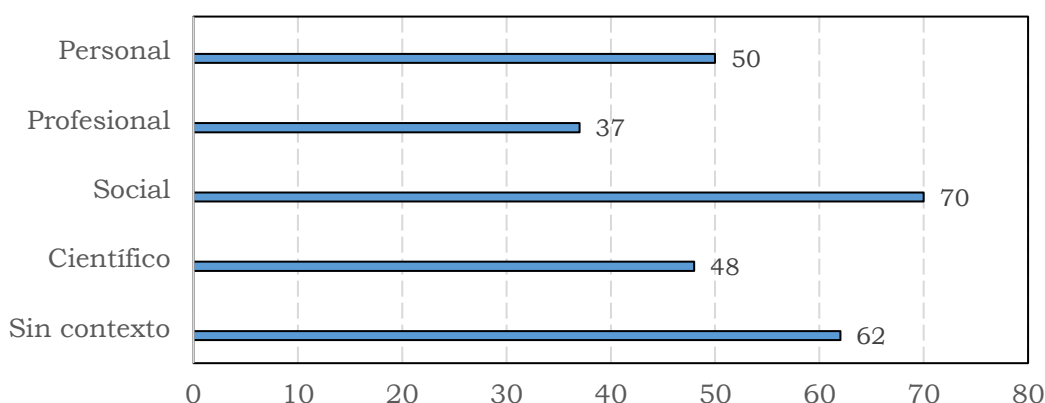


3.9.4. Contexto del problema

Partiendo de la información recogida en la Sección 3.8, se han analizado los contextos en los que se enmarcan los problemas de las pruebas. Con objeto de tener una visión más general sobre la distribución de contextos utilizados se presenta la Figura 3.9.7. Al realizar la clasificación se observa que se tiende a proponer problemas dentro de un contexto social, ya que aproximadamente el 26% de ellos se clasifican como tales. Ello sin duda se debe a la especialidad de Bachillerato a la que se dirigen estas pruebas, que es la de Ciencias Sociales. Sin embargo, se destaca el hecho de que tenga un gran protagonismo los problemas descontextualizados o sin contexto alguno, ascendiendo su

presencia a un 23%. Este resultado es contrario a todas las recomendaciones internacionales sobre la enseñanza de la estadística, hasta el punto de que no se considera verdadera estadística, sino sólo matemáticas el resolver este tipo de tareas. Los demás tipos de contextos se tienen en cuenta en un porcentaje aceptable rodeando el 17%.

Figura 3.9.7. Frecuencia de la variable contexto

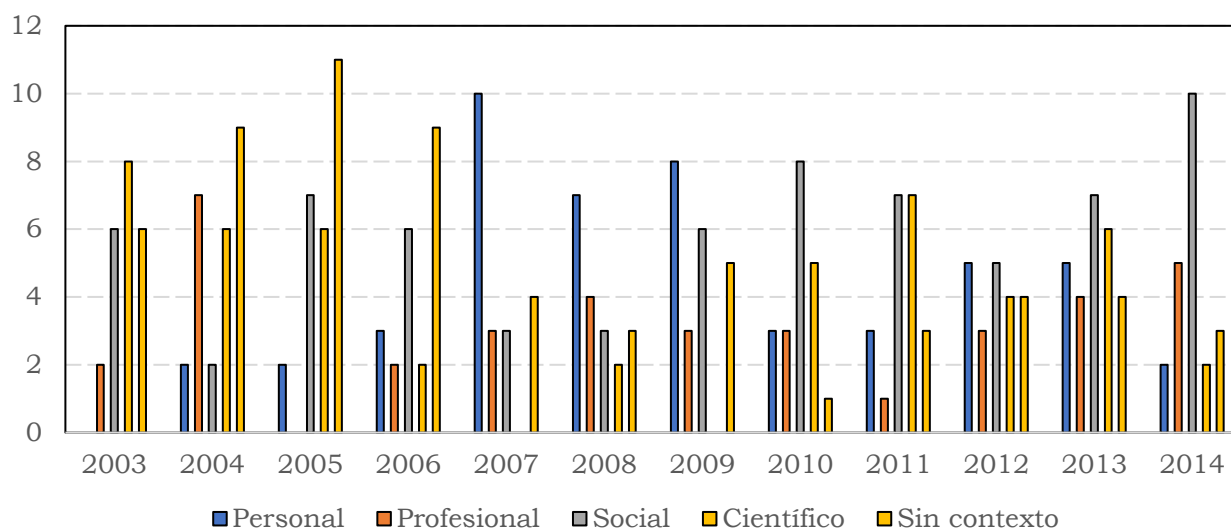


La Figura 3.9.8 muestra la distribución de los problemas analizados según la variable contexto. Se observa una gran presencia de problemas formulados sobre un contexto social y una disminución significativa, en los últimos años, de los problemas sin contexto. Al realizar un análisis comparativo entre el periodo inicial y final del estudio, se observa que ha habido una tendencia a modificar el contexto en el que se desarrolla el problema. Aunque se concluye que aproximadamente el 23% de los problemas analizados carecen de contexto, es necesario señalar que el 76% de ellos han sido propuestos antes de 2010. Este hecho adquiere cierta importancia ya que se tiende a seguir las recomendaciones sobre la enseñanza de la estadística disminuyendo los problemas sin contexto donde el estudiante no comprende el interés de aplicación de las matemáticas.

Se observa gran variabilidad en relación a los contextos, de los 12 años analizados en ocho se han incorporado problemas correspondientes a las cinco categorías descritas para esta variable. Las categorías no incluidas son en 2003 el contexto personal, en 2005 el profesional y en 2007 y 2009 el contexto propiamente

científico. Asimismo, se destaca el año 2007 en el que la mitad de los problemas se enmarca en un contexto personal y en 2014 la mitad a un contexto social.

Figura 3.9.8. *Distribución de la variable contexto en el periodo 2003-2014*



3.9.4.1. Estudio comparativo entre el campo de problema y el contexto

En ciertas ocasiones se puede pensar que el tipo de contexto puede venir relacionado con el tipo de problema que se propone al estudiante en su proceso de enseñanza-aprendizaje. Con objeto de analizar este tema se ha estudiado la relación entre la variable campo de problema (presentada en la Subsección 3.9.1) y la variable contexto (presentado en la Subsección 3.9.4).

La Tabla 3.9.1 recoge el estudio descriptivo de ambas variables. Se observa que la mitad de los problemas descontextualizados o sin contexto corresponden al campo problemas A1 (enumeración de muestras y cálculo de estadísticos de las distribuciones muestrales en poblaciones finitas). Asimismo, se destaca que, de los 40 problemas propuestos de esta categoría, solo el 25% se enmarcan en alguno de los contextos PISA. Este hecho es llamativo ya que, por ejemplo, la enumeración de las distintas muestras, tema de gran calado en la inferencia estadística, únicamente se apoya en la aplicación de reglas matemáticas sin significado. Estos resultados son muy distintos a los obtenidos con los problemas clasificados como A2 (distribución muestral en poblaciones infinitas y cálculo de probabilidades) ya que aproximadamente un tercio corresponden

a un contexto personal, esto se puede pensar que es consecuencia de contextualizar los problemas en los que se solicita principalmente el cálculo de una probabilidad.

Tabla 3.9.1.

Frecuencia de los Problemas Según los Campos de Problema y el Contexto

	Personal	Profesional	Social	Científico	Sin contexto	TOTAL
A1	3	2	5	0	30	40
A2	10	4	6	7	5	32
B1	17	11	23	21	13	85
B2	15	13	23	18	16	85
C	4	7	12	2	0	25
TOTAL	49	37	69	48	64	267

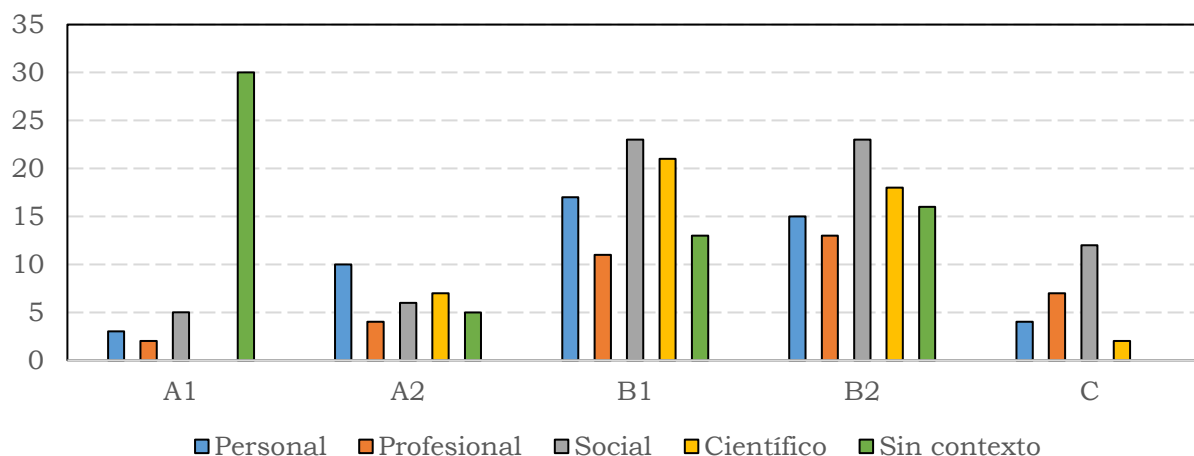
En relación al campo de problemas B (B1, cálculo o interpretación de un intervalo de confianza y B2, relación entre confianza, error de estimación y tamaño muestral) no se detectan diferencias notables en la distribución del tipo de contexto, a excepción del profesional que alcanza en ambos casos un porcentaje del 14% aproximadamente. El resto de contextos tienen una presencia, en media, de 23%. Recordemos que los problemas de la categoría B son los que han tenido, a lo largo del periodo estudiado, una mayor presencia que el resto (64%), sin embargo, se destaca el hecho de que el 17% de ellos no están enmarcados en los contextos contemplados en PISA. Asimismo, centrando la atención en los problemas descontextualizados (64), el 45% de estos corresponden al campo de problemas B.

El análisis de los contextos asociados a los problemas de contrastes de hipótesis llama en gran medida la atención no haber hallado problemas de esta categoría sin contexto. Aproximadamente el 50% se han enfocado en un contexto social, algo propio de la modalidad de Bachillerato que cursan los estudiantes.

La Figura 3.9.10 muestra el estudio de la variable campo de problema en base a las categorías descritas en la Sección 3.8. Se observa la existencia de variabilidad en los contextos empleados en los campos de problemas A2 y el B. Asimismo, se evidencia el uso, e incluso el abuso, que existe en relación a los problemas descontextualizados en el campo A1. Se deduce que los resultados están en la línea de lo expuesto en la Sección 3.9.4 donde se destaca los distintos contextos PISA en el campo de problemas B, en

particular el social con un porcentaje del 67%. Por último, subrayar el no uso de un contexto científico en los problemas de la categoría A1 y que los problemas de la categoría C están incluidos en todas los contextos a excepción de la categoría de problemas sin contexto. La baja presencia de este último tipo de problemas se debe, tal y como se ha mencionado en la Subsección 3.9.1 a su tardía inclusión en las PAU. Sin embargo, aunque diversos estudios han mostrado la dificultad que entraña este campo de problema, el acto de enmarcarlos en un contexto puede suponer facilitar la comprensión sobre el objeto que tiene este tipo de problemas.

Figura 3.9.10. *Distribución de la variable contexto en función del campo de problema*



3.10. CONCLUSIONES DEL ESTUDIO 1

Una vez presentados los resultados del Estudio 1 pasamos a discutir las conclusiones, que se clasifican en los siguientes apartados: idoneidad didáctica de los problemas propuestos en las PAU, conclusiones sobre los objetivos y conclusiones sobre las hipótesis.

3.10.1. Idoneidad didáctica de los problemas de inferencia estadística en las pruebas de acceso a la universidad

Para finalizar el Capítulo 3, hacemos una valoración de algunos componentes de la idoneidad didáctica de las PAU que se pueden deducir del análisis llevado a cabo en las

secciones anteriores. La idoneidad didáctica es introducida por Godino y sus colaboradores para el análisis de procesos de estudio en didáctica de la matemática. Como indican sus autores:

La valoración de la idoneidad de un proceso de instrucción matemática requiere disponer de información detallada de los hechos que ocurren y elementos de referencia que autoricen a emitir los juicios de adaptación, pertinencia o eficacia correspondientes a la dimensión valorada (Godino et al., 2006, p. 4).

Los autores, tal y como se ha indicado en el Capítulo 1, introducen seis componentes en la idoneidad didáctica, que se analizan a continuación en relación a los problemas de inferencia propuestos en las PAU.

Idoneidad epistémica

Mediría la correspondencia entre el significado institucional pretendido (dado por las directrices curriculares) y el evaluado. Es decir, se esperaría que el significado evaluado en las pruebas contuviera una muestra representativa de objetos matemáticos entre los que constituyen el significado pretendido en las directrices curriculares (que a su vez son una muestra representativa del contenido en la propia matemática).

Los resultados del análisis realizado muestran que esta idoneidad es media. Por un lado, cada problema evalúa la comprensión de un gran número de objetos matemáticos, correspondientes a la inferencia estadística, con lo cual, en relación únicamente a la inferencia estadística la idoneidad epistémica sería alta, en general, pero dependerá del problema planteado, pues los problemas relacionados con el muestreo evalúan sólo una parte del significado pretendido en el currículo.

Por otro lado, debemos considerar la representatividad, en relación con todo el contenido curricular del curso citado. Todas las pruebas de los años analizados incluyen un problema de inferencia estadística entre cuatro problemas, al que se añade otro problema de probabilidad condicional o teorema de Bayes (Carretero et al., 2015). Es decir, la cuarta parte de la calificación se otorga por el conocimiento de la inferencia estadística y otra cuarta parte por los temas de probabilidad señalados.

Además, aunque el contenido del currículo incluye otros muchos temas relacionados con inferencia estadística, objetivamente se ve que las pruebas dan mucha mayor importancia al cálculo de intervalos de confianza y la interpretación de los

mismos, que a otros temas de inferencia estadística pues aproximadamente el 64% de los problemas propuestos van relacionados con este tema. De este hecho se deduce que no todos los temas de inferencia se contemplan con el mismo peso en las PAU. Por otro lado, mientras el currículo contempla tres bloques de contenidos -Álgebra, Análisis y Probabilidad y Estadística- la mitad de la prueba está dedicadas a este último bloque de contenidos. En consecuencia, no hay una correspondencia adecuada entre el significado pretendido en el currículo para la matemática en su conjunto y el evaluado en las pruebas de acceso.

Idoneidad cognitiva

Un significado evaluado puede ser representativo del significado pretendido en las directrices curriculares. La idoneidad cognitiva mide el grado en el que los significados evaluados en las pruebas son adecuados a los conocimientos previos de los estudiantes. Aunque no puede ser evaluado directamente en el presente estudio, los resultados de las investigaciones que han analizado el éxito de los estudiantes en las pruebas de acceso a la universidad, detalladas en el Capítulo 2, pueden servir de guía para su análisis.

Pensamos que en este caso esta idoneidad cognitiva, si nos restringimos a los problemas de inferencia estadística es baja, coincidiendo con algunos de los resultados de investigaciones que analizan las calificaciones de los estudiantes (Grau et al., 2002; Murillo, 1997; Ruiz et al., 2013). Así, por ejemplo, Murillo (1997) indica que las pruebas de Matemáticas Aplicadas a las Ciencias Sociales fue la más difícil de todas las asignaturas evaluadas en 29 distritos universitarios. Por ejemplo, García y García (2005) analizan las respuestas dadas por 50 estudiantes a los problemas de inferencia estadística de la convocatoria de junio de 2002 en la Universidad de la Laguna e indica que sólo siete de ellos responden correctamente.

Por otro lado, el análisis de la solución de alguno de estos problemas, revelan que los problemas propuestos son bastante complejos, pues incluyen gran cantidad de objetos matemáticos. Otro indicio de que se quiere aumentar la dificultad es la proporción de problemas de contraste de hipótesis que se plantearon al final del periodo analizado, los cuales requieren el uso de muchos conceptos: población, muestra,

parámetro poblacional, estadístico muestral, distribución del estadístico muestral, hipótesis nula y alternativa, región de aceptación y rechazo, nivel de significación.

Por otra parte, Espinel et al. (2006) seleccionan 272 estudiantes que realizaron la PAU en la Universidad de la Laguna en 2005 y describen los siguientes tipos de errores:

- Confusión entre muestra y población.
- Dificultad para discernir entre parámetro, estimador y estimación.
- Tipificación errónea de la distribución Normal.
- Uso incorrecto de la tabla de probabilidades de la Normal.
- Fallo en la aproximación de la Binomial a la Normal.
- Identificación y formulación confusa de los contrastes unilaterales y bilaterales.
- Intercambio del valor empírico con el valor muestral.
- Asignación incorrecta del nivel de confianza
- Interpretación errónea de las conclusiones en los contrastes

Debemos también señalar que la resolución correcta de este problema no implica que el estudiante comprenda y discrimine todos estos conceptos ni que haya adquirido suficiente razonamiento estadístico, sino que recuerda y sabe aplicar una serie de fórmulas, que quizás no comprenda. Esta enseñanza soslaya también la problemática filosófica asociada y los errores de aplicación de la inferencia estadística. Además, presenta la inferencia frecuencial como una metodología única, ocultando las diferentes aproximaciones y las controversias que dentro de la misma estadística ha tenido la inferencia (Batanero, 2000).

Idoneidad mediacional

Esta idoneidad mide la adecuación de los recursos materiales y temporales que se proporcionan al estudiante para la realización de las pruebas. Fundamentalmente será el tiempo disponible, las tablas de la distribución Normal y las calculadoras. Pensamos que estos recursos son suficientes, sin embargo, es paradójico el hecho de que tanto la estadística como la probabilidad con frecuencia se dejan en los cursos anteriores como último tema y a veces se omiten; el estudiante que se prepara para selectividad tendrá

que hacer un esfuerzo notable para adquirir suficiente competencia y comprensión para resolver los problemas propuestos en las PAU.

También es notable que los problemas estén pensados para ser resueltos a mano y en todo caso, con la tabla de la distribución Normal o la calculadora. Actualmente la práctica de la estadística se simplifica notablemente mediante el uso de ordenadores. Pensamos que este sería un punto a tener en cuenta para modificar en futuras pruebas.

Idoneidad emocional

Mide el interés o motivación que puede tener el estudiante con el objeto de estudio. No podemos deducirlo directamente, pues se requeriría una encuesta entre los estudiantes. En nuestro caso, al analizar los contextos que se presentan en los problemas, se ha observado un predominio del contexto social que es el que en principio debe tener más interés para los estudiantes de esta rama de Bachillerato. Llamamos la atención al 21% de problemas descontextualizados donde el estudiante no puede ver el interés o la aplicación que pueda tener la estadística en su futura vida profesional.

Idoneidad ecológica

En este caso, se pretende medir el grado en que el proceso de estudio se ajusta al proyecto educativo de la institución y a los condicionamientos del entorno en que se desarrolla. Pensamos que esta idoneidad si es tratada correctamente puesto que las PAU cumplen su función de evaluar los conocimientos y competencias de los estudiantes antes de su entrada a la universidad. Aunque sin duda el proceso se puede mejorar y existen propuestas de que cada facultad proponga sus propias pruebas, este es un punto que necesita mucha reflexión y discusión, pues no todos los estudiantes en principio tienen claro los estudios que desean cursar en su nueva etapa universitaria.

En resumen, nuestro análisis indica que algunos componentes de la idoneidad didáctica son claramente mejorables, en particular, la idoneidad epistémica y cognitiva, por lo que debería ser tenida en cuenta por los diseñadores de las mismas en las sucesivas ediciones o en pruebas de evaluación alternativas que se propongan en el futuro.

3.10.2. Conclusiones sobre los objetivos

Se comienza presentando los objetivos específicos, que se describieron en el apartado 3.2 y las conclusiones obtenidas respecto a cada uno de ellos:

OE2.1 Analizar los campos de problemas sobre inferencia estadística propuesto en PAU, identificando los objetos matemáticos requeridos en su solución

El análisis cualitativo y cuantitativo de los problemas propuestos en las PAU a lo largo de 12 años en Andalucía ha permitido identificar cinco grupos de campos de problemas. En base al enfoque ontosemiótico, se han resuelto cada uno de los problemas propuestos en las pruebas PAU para Matemáticas Aplicadas a las Ciencias Sociales II durante el periodo 2003-2014, y analizado las prácticas matemáticas requeridas en su solución. A partir de este análisis se han identificado los objetos relacionados con la inferencia estadística. Todo ello para determinar el significado institucional evaluado en las pruebas y compararlo con el significado institucional pretendido en las orientaciones curriculares.

Se identificaron cinco campos de problemas diferenciados: A1) Enumeración de muestras y cálculo de estadísticos de las distribuciones muestrales en poblaciones finita; A2) Distribución muestral en poblaciones infinitas y cálculo de probabilidades; B1) Cálculo o interpretación de un intervalo de confianza, B2) Relación entre confianza, error de estimación y tamaño muestral y C) Contraste de hipótesis. El análisis semiótico de un ejemplo de cada campo ha permitido, describir las prácticas matemáticas requeridas en su resolución. A partir de dichas prácticas se ha podido formar las configuraciones epistémicas requeridas en la solución de cada campo de problema y mostrar sus diferencias y similitudes. Los resultados muestran la alta complejidad de los problemas de inferencia propuestos en las pruebas, especialmente los referidos a contraste de hipótesis e intervalos de confianza. La comparación con los contenidos curriculares y los estándares previstos para la inferencia estadística, tanto en la anterior legislación como en la actual, muestra un buen ajuste entre el significado institucional pretendido en los currículos y el evaluado en las PAU, con algunas matizaciones que se describen en la discusión de las hipótesis. También se ha constatado que en algunos problemas se incluye más de un campo de problemas, lo que lleva a una evaluación más completa de los contenidos citados.

OE2.2. Identificar las variables que definen los problemas de inferencia estadística propuestos en las PAU y analizarlas estadísticamente.

Para conseguir este objetivo, se parte de la investigación previa sobre comprensión de la inferencia estadística resumida en el Capítulo 2. Esta investigación muestra numerosas dificultades asociadas a los diferentes conceptos, propiedades y procedimientos ligados a los problemas de inferencia estadística. Indirectamente dicha investigación nos informa sobre algunas variables de los problemas de inferencia que inciden directamente en su dificultad.

Por otro lado, partiendo de la información ofrecida por el objetivo específico anterior, es decir, los principales campos de problemas, se identifican la distribución de partida, el parámetro a estimar y el contexto, como variables objeto de estudio, para caracterizar los problemas propuestos.

Una primera revisión de los problemas permitió identificar categorías para cada una de las variables anteriores. Posteriormente, se codifican los datos para presentar la distribución de cada una de las variables globalmente y por año, proporcionando un análisis detallado de estos resultados.

OE2.3. Analizar la idoneidad didáctica de los problemas de inferencia propuestos en las pruebas PAU.

Como síntesis y conclusión de los anteriores análisis se ha valorado en la Subsección 3.10.1 la idoneidad didáctica de los problemas analizados, siguiendo los indicadores de Godino (2009, 2013) y teniendo en cuenta los análisis realizados en el Capítulo 2 y en este capítulo.

3.10.2. Conclusiones sobre las hipótesis

Analizados los objetivos pasamos a discutir las hipótesis que se plantearon al comienzo del estudio y que son las siguientes:

H2.1. Se espera encontrar una alta importancia de la inferencia estadística en las PAU, reflejada a través de un elevado número de problemas relacionados con la inferencia estadística.

Nos basamos para formular esta hipótesis en un análisis preliminar publicado por Díaz et al. (2014). Esta hipótesis se ha confirmado, pues la alta importancia dada a la inferencia estadística se pone de manifiesto en el hecho de que de los cuatro problemas que ha de resolver el estudiante uno de ellos corresponde a este contenido. Siendo la inferencia estadística tan sólo una parte del Bloque de Estadística y Probabilidad y añadido a la presencia de otros bloques, es claro que la inferencia estadística recibe en estas pruebas de evaluación un peso mayor al fijado en el currículo.

H2.2. Además, se observará un buen ajuste entre los contenidos de inferencia estadística fijados en el currículo y los considerados en las pruebas.

Por otro lado, el análisis semiótico de la solución de un problema de cada uno de los campos identificados ayudó a identificar la multitud de objetos matemáticos que indirectamente son sometidos a evaluación. Al comparar estas configuraciones de objetos con los contenidos y estándares curriculares se encontró en general un buen ajuste del significado pretendido y el evaluado, con ligeras excepciones.

Todos los contenidos fijados en el significado previsto de la inferencia estadística en las directrices curriculares citadas se evalúan en algunos de los campos de problemas planteados, aunque no con la misma intensidad. Por otro lado, los problemas relacionados con la elección de las muestras sólo se evalúan en los campos de problemas relacionados con ellas, es decir, sobre tipos de muestreo. El conocimiento sobre intervalo de confianza o contraste de hipótesis, sus propiedades y los procedimientos para llevarlos a cabo se evalúan en los campos de problemas B o C respectivamente y el conocimiento de la distribución muestral en todos los campos propuestos.

También se observa que, tanto el cálculo de estimadores puntuales (media, varianza, desviación típica y proporción poblacionales), como el cálculo de probabilidades asociadas a la distribución de la media muestral y de la proporción muestral se recogen en todos los campos de problemas analizados. Nótese que el resto de estándares de aprendizaje son específicos del intervalo de confianza o el contraste de hipótesis. Por último, se debe señalar que, aunque la representatividad de la muestra es un elemento de gran importancia dentro de la inferencia estadística, ninguno de los campos de problemas recoge el estándar de la valoración de la representatividad de la muestra.

H2.3. *Aunque en el currículo que han cursado los estudiantes en los pasados años (MEC, 2007) se sugiere estudiar la inferencia estadística tanto para una como para dos muestras y en relación a la media y proporción, pensamos que la mayoría de problemas se centrarán tan sólo en la inferencia sobre la media de la muestra.*

Esta hipótesis se confirma pues no se ha encontrado problemas relacionados con el contraste de hipótesis o el intervalo de confianza en dos poblaciones. Los problemas relativos a la comparación de dos muestras, tanto para la media como para la proporción han sido suprimidos del currículo en los últimos documentos oficiales (MECD, 2015). En este sentido, aunque las pruebas analizadas corresponden al periodo anterior a la aplicación de las nuevas directrices, los problemas propuestos guardan mayor relación con estos currículos que con los anteriores.

H2.4. *Al estudiar el contexto de los problemas, pensamos que se encontrarán una proporción grande de problemas descontextualizados sobre los problemas de probabilidad y pensamos se repetirá en nuestro estudio.*

Esta hipótesis fue confirmada en el estudio que realiza Carretero (2014) y se ha constatado en nuestro estudio pues se presenta una proporción importante de problemas sin ningún contexto significativo. Sin embargo, esta proporción es menor comparada con la obtenida en el estudio de Carretero y, por otro lado, se han encontrado en los problemas todos los contextos considerados en las pruebas PISA.

La proporción de problemas descontextualizados ha disminuido gradualmente en las últimas pruebas analizadas en las que aparecen raramente. Por tanto, entendemos que se trata actualmente de seguir las directrices sobre la importancia del contexto tanto de las pruebas PISA, como dentro de la misma estadística.

EVALUACIÓN DE CONOCIMIENTOS DE FUTUROS PROFESORES

Índice de Contenidos

- 4.1. Introducción
 - 4.2. Objetivos del Estudio 2
 - 4.3. Hipótesis o expectativas iniciales
 - 4.4. Metodología
 - 4.4.1. Muestra y contexto
 - 4.4.2. Cuestionario propuesto
 - 4.4.3. Método de análisis de los datos
 - 4.5. Conocimiento matemático sobre el contraste de hipótesis
 - 4.5.1. Planteamiento de las hipótesis
 - 4.5.2. Elección del tipo de contraste
 - 4.5.3. Desarrollo del procedimiento
 - 4.5.4. Determinación del p -valor
 - 4.5.5. Interpretación de los resultados
 - 4.5.6. Síntesis de conocimientos común sobre el contraste de hipótesis
 - 4.5.7. Conclusiones sobre el conocimiento del contraste de hipótesis
 - 4.6. Conocimiento matemático sobre el intervalo de confianza
 - 4.6.1. Planteamiento del intervalo de confianza
 - 4.6.2. Cálculo del intervalo
 - 4.6.3. Interpretación del intervalo
 - 4.6.4. Definición
 - 4.6.5. Propiedades que se recuerdan
 - 4.6.6. Síntesis de conocimientos sobre el intervalo de confianza
 - 4.6.7. Conclusiones sobre el conocimiento común del intervalo de confianza
 - 4.7. Conocimientos didácticos sobre inferencia estadística
 - 4.7.1. Errores en el planteamiento de las hipótesis
 - 4.7.2. Errores en la selección del contraste de hipótesis e intervalo de confianza
 - 4.7.3. Errores conceptuales
 - 4.7.4. Errores procedimentales
 - 4.7.5. Errores de interpretación
 - 4.7.6. Formas de explicar el intervalo de confianza en el aula
 - 4.7.7. Síntesis de conocimientos didácticos
 - 4.7.8. Conclusiones sobre el conocimiento didáctico sobre inferencia estadística
 - 4.8. Conclusiones del Estudio 2
 - 4.8.1. Conclusiones sobre los objetivos
 - 4.8.2. Conclusiones sobre las hipótesis
-

4.1. INTRODUCCIÓN

En este capítulo se presenta el estudio de evaluación de los conocimientos que muestra un grupo de futuros profesores de Educación Secundaria Obligatoria y Bachillerato sobre contenidos de inferencia estadística. En concreto, en este estudio se evalúan algunos elementos del conocimiento matemático común y avanzado de los futuros profesores, así como de la faceta cognitiva de su conocimiento didáctico-matemático (Godino et al., 2017; Pino-Fan y Godino, 2015) respecto a dos procedimientos de inferencia: el contraste de hipótesis y el intervalo de confianza. Se incluye también una pregunta sobre cómo explicarían el intervalo de confianza que formaría parte de la faceta mediacional de su conocimiento.

A lo largo del capítulo se exponen en primer lugar los objetivos que sustentan el estudio junto a las hipótesis iniciales sobre los resultados que se prevén obtener, que se apoyan en la revisión bibliográfica minuciosa presentada en el Capítulo 2. En las secciones siguientes se detalla la muestra formada por futuros profesores y el cuestionario empleado para analizar algunos elementos del conocimiento, tanto matemático como didáctico, que poseen dichos participantes en el tema de inferencia estadística.

Seguidamente, se presentan los resultados de los análisis realizados de las respuestas a cada una de las preguntas planteadas que se acompañan con una síntesis de sus conocimientos matemáticos y didácticos (respecto a las preguntas planteadas). Para finalizar se destacan las conclusiones principales del estudio en función a los objetivos e hipótesis que se describieron en una primera etapa. Somos conscientes de las limitaciones del estudio, pues sólo se evalúan algunos elementos del amplio conocimiento didáctico-matemático, en sus diferentes componentes que debieran tener los futuros profesores. Esta limitación se asume de antemano, pues se ha preferido profundizar en el análisis de las respuestas, limitando el número de preguntas.

No obstante, dada la escasez de investigaciones sobre el conocimiento didáctico-matemático de los futuros profesores sobre inferencia estadística, la información proporcionada será de utilidad a investigadores y formadores de profesores. Parte de los resultados presentados en este capítulo han dado origen a artículos como Batanero y López-Martín (en prensa), Batanero, López-Martín, Gea y Arteaga (2019) y López-Martín,

Batanero y Gea (2019), así como a las presentaciones en congresos listadas en el Apéndice.

4.2. OBJETIVOS DEL ESTUDIO 2

A lo largo de esta sección se detalla el objetivo principal asociado a este estudio, desglosándolo a su vez en tres objetivos parciales. Los objetivos de este estudio de evaluación desarrollan el tercer objetivo general de la investigación, que era el siguiente:

OE3. Evaluar algunos elementos del conocimiento didáctico-matemático sobre la inferencia en futuros profesores de Educación Secundaria Obligatoria y Bachillerato.

Este objetivo se justifica por la escasez de investigaciones relacionadas con el conocimiento del profesor en este tema, que se mostró al presentar en el Capítulo 2 el estado de la cuestión. Una segunda razón que avala el estudio es la necesidad de que estos profesores tengan una formación suficiente sobre un tema que deben enseñar, ya que está incluido en las directrices curriculares (analizadas en el Capítulo 1) y además suele ser evaluado en las pruebas de acceso a la universidad, como se ha mostrado en el estudio empírico de los problemas propuestos en dichas pruebas presentado en el Capítulo 3.

Como se indicó al describir este objetivo en el Capítulo 1, para cumplirlo ha sido necesario preparar un cuestionario dirigido a futuros profesores de los citados niveles educativos con el fin de recoger datos para su posterior análisis. Esta recogida se ha llevado a cabo como parte de un taller formativo dirigido a desarrollar el conocimiento didáctico-matemático (Godino et al., 2017; Pino-Fan y Godino, 2015) sobre inferencia estadística que se llevó a cabo en el marco del Máster de Formación del Profesorado de Educación Secundaria Obligatoria y Bachillerato, especialidad de Matemáticas.

El estudio llevado a cabo ha permitido dividir el objetivo principal, OE4, en tres objetivos parciales, que son descritos a continuación:

OE3.1. Evaluar el conocimiento matemático común y avanzado sobre el contraste de hipótesis en futuros profesores de Educación Secundaria Obligatoria y Bachillerato.

Este conocimiento común se ha deducido de los contenidos sobre el contraste de hipótesis incluidos en el Real Decreto de Enseñanzas Mínimas (MEC, 2007), y del análisis de los problemas de contraste de hipótesis propuestos en las PAU, es decir, del análisis semiótico de los problemas tipo C descritos en el Estudio 1 y resumido en la Tabla 3.5.6. En base al análisis, y teniendo en cuenta los principales objetos matemáticos identificados en el mismo, se desea evaluar: el conocimiento matemático avanzado que poseen los futuros profesores en relación al planteamiento que realizan a la hora de determinar las hipótesis del contraste, la elección del contraste (bilateral o unilateral), el desarrollo del procedimiento seleccionado según la metodología seguida el cálculo del p -valor y la interpretación de los resultados obtenidos.

El contraste de hipótesis ha sido uno de los temas de Bachillerato en la LOE (MEC, 2007) y fueron suprimidos en la LOMCE (MECD, 2015b). En el periodo 2007-2014 fue uno de los problemas frecuentes en las PAU, como se ha analizado en el Capítulo 3. Por tanto, el conocimiento evaluado sería parte del conocimiento matemático común del profesor en los periodos en que el currículo incluye la enseñanza del contraste de hipótesis, como ocurría en las anteriores directrices curriculares. En la actualidad, dada la supresión de dicho contenido, se puede considerar que se trata de evaluar el conocimiento avanzado del contenido, ya que corresponde a temas que se enseñarían en cursos posteriores a Bachillerato.

Los conocimientos concretos evaluados se han elegido por ser aquellos en los cuáles se han descrito un mayor número de errores en los estudiantes (véase Capítulo 2).

OE3.2. Evaluar el conocimiento matemático común sobre el intervalo de confianza en futuros profesores de Educación Secundaria Obligatoria y Bachillerato.

En concreto, se desea evaluar el conocimiento que tienen los futuros profesores sobre el planteamiento del intervalo de confianza, el desarrollo del procedimiento que permite la obtención de los valores numéricos de los extremos del intervalo y la interpretación de los resultados. Asimismo, se pretende evaluar la definición de dicho concepto y las propiedades que recuerdan los futuros profesores sobre el este.

El intervalo de confianza ha sido y sigue siendo uno de los temas trabajados en el segundo curso de Bachillerato de la modalidad de Humanidades y Ciencias Sociales.

Desde su introducción en las normativas curriculares los problemas de intervalos de confianza han adquirido gran relevancia dada la frecuencia de aparición en las PAU (véase Capítulo 3). Luego, con la presentación de este objetivo, se pretende evaluar un componente del conocimiento común del contenido por parte de los futuros responsables de su enseñanza en dicho nivel educativo. Es además uno de los procedimientos que se recomiendan para complementar los contrastes de hipótesis en la investigación (APA, 2010). Por otro lado, la revisión bibliográfica realizada en el Capítulo 2 pone de manifiesto los numerosos errores de interpretación del intervalo de confianza, que podrían compartir los futuros profesores.

Los conocimientos concretos evaluados se han elegido por ser aquellos en los que se han descrito un mayor número de errores por parte de los estudiantes (ver Capítulo 2). Adicionalmente, se desea evaluar si los futuros profesores son conscientes y conocen los posibles errores y dificultades de sus estudiantes con el tema. De ello se deduce un nuevo objetivo.

OE3.3. Evaluar algunos elementos de la faceta cognitiva del conocimiento didáctico-matemático de los futuros profesores de Educación Secundaria Obligatoria y Bachillerato. En concreto, se desea evaluar el conocimiento que los futuros profesores de Educación Secundaria Obligatoria y Bachillerato tienen de los posibles errores de los estudiantes en el contraste de hipótesis e intervalo de confianza.

En concreto, se desea evaluar su conocimiento de los errores descritos en la investigación previa sobre los contrastes de hipótesis y el intervalo de confianza. Además, se pretende relacionar los errores que cometen ellos mismos en estos procedimientos con los errores que esperan encontrar en sus estudiantes. Estos conocimientos serían parte de la componente cognitiva del conocimiento del profesor sobre inferencia estadística (Pino-Fan y Godino, 2015).

La revisión bibliográfica realizada en el Capítulo 2 ha puesto de manifiesto la gran variedad de errores existentes a la hora de abordar un problema de contraste de hipótesis y/o de intervalo de confianza (por ejemplo, nivel de significación, p -valor, establecimiento de las hipótesis, etc.). Un conocimiento de los errores por parte del profesorado encargado de transmitir dichos contenidos garantiza una mejora en el aprendizaje del alumnado. Reconocer la posibilidad de cometer un error contribuye a

erradicar dicho error.

Con objeto de analizar el conocimiento que posee los futuros profesores en relación a los errores existentes en la inferencia estadística, se le solicita que listen el máximo número de errores que podrían cometer los estudiantes cuando elaboran un problema similar al realizado por ellos mismos. Como complemento se añade junto al ítem 2 una pregunta sobre la forma en que explicarían el intervalo de confianza a sus estudiantes, permitiendo evaluar una pequeña parte de la componente mediacional del conocimiento didáctico-matemático de estos profesores.

4.3. HIPÓTESIS O EXPECTATIVAS INICIALES

Se formulan algunas hipótesis o expectativas iniciales que se esperan encontrar en la elaboración del presente estudio. De los objetivos señalados, así como del estudio de los antecedentes, se deducen algunas hipótesis o expectativas iniciales que podríamos analizar a partir de los resultados de la evaluación y que son las siguientes:

H3.1. Al evaluar los conocimientos de los futuros profesores de Educación Secundaria Obligatoria y Bachillerato se prevé encontrar algunas dificultades, similares a las descritas en estudiantes en conceptos estadísticos requeridos para el trabajo con la inferencia estadística que deben enseñar, esto es, en su conocimiento común del contenido.

Esta hipótesis se deduce del hecho de que en estudios previos con licenciados universitarios que han tenido una formación similar a la de los futuros profesores participantes en nuestro estudio, se han observado numerosos errores, tanto en el contraste de hipótesis como en el intervalo de confianza.

En concreto, se espera encontrar errores en el planteamiento de las hipótesis en un contraste (Batanero et al., 2012; Cañadas et al., 2012; Vallecillos, 1994), y en la determinación de la región crítica (Batanero, Vera y Díaz-Batanero, 2012; Vallecillos, 1999). Asimismo, se espera encontrar ciertas limitaciones a la hora de interpretar correctamente el nivel de confianza y/o nivel de significación (Batanero, 2000; Birnbaum 1982; Falk, 1986; Vallecillos; 1994) y diferenciar los estadísticos muestrales de los parámetros poblacionales (Harradine et al., 2011; Olivo, 2008). Igualmente pensamos se producirán errores en la identificación del modelo probabilístico de la

variable media muestral, que usualmente será la distribución normal. Los estudiantes podrían no identificar dicha distribución o bien cometer un error al determinar sus parámetros lo cual provocará errores en el proceso de tipificación de la variable (Espinel et al., 2007). Respecto al intervalo de confianza se espera la interpretación incorrecta de sus extremos, asumiéndolos como fijos (Behar, 2001; Olivo, 2008) o falta de comprensión de la relación del ancho del intervalo con diferentes factores (Cumming y Fidler, 2005).

H3.2. Se espera también encontrar algunas dificultades al emplear procedimientos estadísticos básicos, requeridos en la aplicación de la inferencia, es decir en la parte procedimental de su conocimiento común de la inferencia estadística.

Esta conjetura se fundamenta en los errores encontrados en la literatura sobre el uso de las tablas estadísticas y la tipificación de variables aleatorias continuas (Alvarado, 2007). Además, el análisis de las investigaciones relacionadas con el tema de estudio ha permitido verificar que se tiende a describir la desviación típica de la variable aleatoria media muestral igual que la poblacional obviando que ésta debe ir dividida por la raíz cuadrada del tamaño muestral, es decir, no se realiza una aplicación adecuada del teorema central del límite (Alvarado, 2007; Batanero et al., 2001; Vallecillos, 1994). Por otro lado, se espera encontrar errores en la obtención de los cuantiles que determinan las regiones de rechazo en un contraste de hipótesis o que son empleados para la obtención de los intervalos de confianza, debido a la dificultad en el manejo de desigualdades o a la confusión entre región crítica y de aceptación (Chance et al., 2004; Espinel et al., 2007; Inzunza, 2006).

H3.3. Los participantes tendrán más dificultad en la interpretación de los resultados de la inferencia estadística que en el planteamiento de los procedimientos o en la realización de los cálculos asociados.

La investigación llevada a cabo por Espinel et al. (2007) sobre el análisis de errores más frecuentes que cometen los estudiantes cuando abordan las cuestiones de inferencia estadística incluidas en las PAU de la Universidad de La Laguna, muestra, entre otras, las dificultades existentes a la hora de dar una interpretación sobre las conclusiones que se desprende de la realización de un contraste. Estos estudios, junto a los resultados obtenidos por diversos autores, ponen de manifiesto la dificultad que tienen tanto los estudiantes como investigadores cuando deben interpretar la

información facilitada por un contraste o por un intervalo de confianza (Batanero, 2000; Cumming et al., 2004; Fidler y Cumming, 2005; Vallecillos, 1994).

Estos resultados hacen pensar que los futuros profesores puedan tener ciertas dificultades en la interpretación y contextualización de los resultados obtenidos. Esta hipótesis adquiere un gran interés, ya que la realización de este tipo de actividades, sobre todo en temas relacionados con investigaciones, tiene como objeto concluir sobre una/s característica/s determinada/s de una población no observable a partir de la información de una muestra extraída de la misma. Luego una interpretación adecuada junto con su contextualización conlleva una correcta toma de decisiones sobre la población bajo estudio.

H3.4. Los futuros profesores de Educación Secundaria Obligatoria y Bachillerato desconocen parte de los errores de aplicación e interpretación de la inferencia estadística que se han descrito en la investigación previa, mostrando carencias en la faceta cognitiva de su conocimiento didáctico-matemático.

Esta hipótesis se deduce de las investigaciones que se han llevado a cabo dentro del área de la inferencia estadística donde se muestran los diversos errores que suelen estar presentes tanto en estudiantes como en investigadores. Es de suponer, por tanto, que los sujetos no son conscientes de los errores. Además, en investigaciones como la de Krauss y Wassner (2002) son los mismos profesores los que cometen estos errores.

La revisión bibliográfica llevada a cabo en el Capítulo 2 ha permitido ver que errores que se han analizado con anterioridad se siguen cometiendo en la actualidad tanto por estudiantes como por personal docente e investigador. Este hecho, nos hace pensar que la repetición de estos tipos de errores a lo largo del tiempo viene provocada, entre otros, por el desconocimiento existente en la comunidad docente en relación al tema en estudio. Por tal motivo, pensamos que los futuros profesores no reconocen los errores más destacados que suelen ser cometidos por los estudiantes.

4.4. METODOLOGÍA

A lo largo de esta sección se describe con detalle la metodología empleada en el Estudio 2, incluyendo sus características generales, la muestra y el contexto donde se lleva a cabo. Se muestra también el cuestionario, de elaboración propia, a partir del cual se

analizan algunos elementos del conocimiento didáctico-matemático de los futuros profesores de Educación Secundaria Obligatoria y Bachillerato con los que se ha contado. Igualmente se describe el método de análisis de los datos recogidos a partir del cuestionario.

Se trata de una investigación predominantemente cualitativa, pues se trabaja con respuestas abiertas de los futuros profesores a ciertas tareas y, por tanto, los datos son de tipo cualitativo. Es inductiva, pues las categorías en que se clasifican las variables no están fijadas a priori, sino que se obtienen mediante un proceso cíclico donde son depuradas. Por otro lado, se trata de un estudio a pequeña escala y de naturaleza flexible. Se complementa con algunos análisis estadísticos elementales, principalmente tablas de frecuencias y algunas gráficas.

Según Bisquerra y Alzina (2004), la investigación cualitativa se caracteriza por la recogida de datos en ambiente natural; en nuestro caso se recogieron como parte de un taller formativo, colaborando la profesora del curso con la investigadora y contando con una observadora participante del proceso que también ayudó, cuando fue necesario, tanto en la recogida de datos como en su interpretación. Otra característica descrita por estos autores es la preferencia por los procesos, frente a los productos. Es por ello por lo que nos interesaremos por las estrategias de resolución y no solamente si la respuesta es correcta o no. Se basa en el análisis inductivo de los datos, que se ha utilizado en nuestro caso.

El estudio de evaluación realizado se considera exploratorio (Hernández et al., 1994) pues analizamos un problema poco estudiado hasta la fecha. Sin embargo, se cuenta con estudios realizados con estudiantes universitarios, lo que nos ha permitido plantear hipótesis en el sentido de expectativas y no como hipótesis estadísticas.

4.4.1. Muestra y contexto

Para llevar a cabo el presente estudio, se contó con la colaboración de 73 estudiantes del Máster Universitario en Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas, (MAES), especialidad de Matemática. A pesar de que la especialidad corresponde a Matemáticas, solo algo más de la mitad poseen el Grado de Matemáticas o Estadística (56%), correspondiendo el

resto a otras ramas de ciencias (ingeniería, arquitectura, entre otras). No obstante, todos los participantes reconocen haber cursado, en su etapa universitaria, una o más asignaturas de estadística y el 57% señalan tener experiencia de enseñanza de la materia de Matemáticas en diferentes niveles educativos. Los estudiantes estuvieron divididos en dos grupos de aproximadamente el mismo tamaño, que trabajaron con el mismo profesor y método. La recogida de datos se lleva a cabo, a lo largo del curso 2017-2018, en la Universidad de Granada. En lo que sigue, se describen las características generales de este Máster y de la asignatura en la que se toman los datos.

El MAES es impartido en más de 40 universidades españolas y está orientado a formar a profesionales de la docencia en niveles de Educación Secundaria, Bachillerato, Formación Profesional y Enseñanzas de Idiomas tanto en centros públicos como privados y es obligatorio haberlo cursado según establece Ley Orgánica de Educación 2/2006 de 24 de mayo de 2006 y en la regulación establecida para estos máster en la Orden 3858/2007 de 27 de diciembre. La organización de las enseñanzas del citado Máster es común en las distintas universidades y especialidades (matemáticas, lengua, etc.), y se estructura, como mínimo, en los tres módulos: 1) El módulo genérico que abarca las asignaturas de tipo psicológico o pedagógico general, así como la legislación educativa y aspectos de la organización de los centros escolares; 2) El módulo específico centrado en el desarrollo de las competencias relacionadas con las diferentes disciplinas o especialidades docentes; 3) El módulo de Prácticum, que corresponde al periodo de prácticas en instituciones educativas. Para finalizar los estudios de Máster, se presenta un Trabajo Fin de Máster relacionado con la formación adquirida.

El módulo específico de Matemáticas se compone de tres materias: 1) Complementos de formación disciplinar en el que se refuerza al estudiante en temas sobre historia y filosofía de las matemáticas o bien temas poco conocidos por el estudiante; 2) Aprendizaje y enseñanza de las materias, centrada en bloques de contenido tales como competencia matemática y análisis curricular, conocimiento profesional del profesor de matemáticas, aprendizaje matemático y características del razonamiento matemático de los estudiantes de Educación Secundaria Obligatoria y Bachillerato, así como diseño y análisis de unidades didácticas y programaciones de aula; 3) Innovación docente e investigación educativa, en el que se abordan los recursos didácticos, tecnología educativa y métodos de enseñanza de las Matemáticas.

El estudio que se desarrolla a lo largo del presente capítulo se lleva a cabo en el marco de la asignatura Innovación Didáctica e Iniciación a la Investigación en Matemáticas tiene un carácter esencialmente práctico y se desarrolla presencialmente en 15 sesiones de dos horas y media de duración, dividiéndose en dos períodos de docencia. El primer período (5 sesiones) es común a todas las especialidades, mientras el segundo (10 sesiones) se emplea para la formación específica a la especialidad de matemáticas. Los contenidos específicos fijados en la guía docente de la asignatura son los siguientes:

1. El proceso de enseñanza y aprendizaje de las matemáticas. Criterios de idoneidad.
2. Investigación en educación matemática. Problemas y líneas. Criterios de calidad. Análisis de ejemplos.
3. Innovación en educación matemática. Fundamentación y finalidad. Buenas prácticas en la enseñanza. Análisis de ejemplos.
4. Recursos para la innovación e investigación: documentos, materiales, tecnología, metodología, interdisciplinariedad.
5. La difusión y aplicación en el aula de los resultados de la investigación e innovación. Fuentes de información.
6. El profesor como investigador e innovador. Análisis crítico de la práctica docente.

La asignatura está basada en la realización de talleres en donde los estudiantes trabajan en ocasiones individualmente y en otras en pareja, para luego discutir en grupo las respuestas a las tareas, que se orientan a desarrollar en el estudiante la capacidad de análisis didáctico. En este capítulo se analizan los datos obtenidos en dos actividades de análisis didáctico llevadas a cabo en dos sesiones prácticas, de aproximadamente dos horas de duración. Durante el periodo, se siguió el ciclo formativo propuesto por Godino et al. (2008) con objeto, no sólo de evaluar, sino también a mejorar el conocimiento del profesor. Para ello, en una primera etapa, y en base a un proyecto estadístico que puede ser empleado en Bachillerato, se trabaja con los futuros para posteriormente continuar con una fase de análisis didáctico basado en su experiencia personal con dicho proyecto.

En la primera fase del taller, dirigida a desarrollar y evaluar el conocimiento del contenido de inferencia de los participantes, se resolvió un problema de contraste de hipótesis y otro de intervalo de confianza. Los estudiantes dispusieron de una hora para resolver cada problema y, en la segunda parte de la sesión, se dedicó a corregir y discutir con los estudiantes las soluciones, una vez recogidos los cuestionarios. En la segunda sesión, dedicada a desarrollar su conocimiento didáctico, se les pidió resumir las dificultades que a su juicio podrían tener los estudiantes de bachillerato con los problemas propuestos y describir cómo explicarían el intervalo de confianza a sus futuros alumnos.

4.4.2. Cuestionario propuesto

La elaboración del cuestionario es una adaptación de la investigación realizada por Gea (2014) en el que la autora trabajó con un proyecto estadístico, alrededor de una pregunta de investigación, ¿Cuáles son los factores que más influyen en la esperanza de vida al nacer en un país? Esta pregunta permitió plantear diferentes tareas relacionadas con la inferencia estadística y no incluidas en la versión del proyecto que utilizó dicha autora.

Se eligió el proyecto como modalidad de trabajo en aula con objeto de contextualizar el tema en cuestión permitiendo a los participantes pasar por las distintas fases de una investigación estadística: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado. El uso de este tipo de recurso didáctico se fundamenta, por un lado, en que fomenta el desarrollo del modelo de razonamiento estadístico propuesto por Wild y Pfannkuch (1999) y, además, es reconocido como una herramienta de interés en las orientaciones metodológicas del currículo (MECD, 2015b). Concretamente los datos fueron extraídos del servidor de la Naciones Unidas (<http://hdr.undp.org/es/data>) aportando una gran variedad de variables agrupada en categorías (por ejemplo, salud, género, igualdad, economía, educación, etc.) de un elevado número de países (en nuestro caso se seleccionaron 194 países). Para el proyecto se seleccionaron nueve variables del último año disponible (2009), siendo la esperanza de vida al nacer la principal variable dependiente para el estudio de la inferencia estadística en el proyecto. El resto de variables se han elegido para tener en

cuenta diversas intensidades, signos y tipos de relación con la esperanza de vida, pero no las tenemos en cuenta en el Estudio 2.

Para llevar a cabo el proyecto, inicialmente, el formador de profesores analizó junto a los participantes el significado de cada una de las diferentes variables seleccionadas y la forma en que se recogían los datos. Una vez comprendido su significado, se fue proporcionando, sucesivamente, cada una de las tareas del cuestionario que se analizan a continuación, dejando un tiempo suficiente. Tras haber finalizado la primera tarea, se debatieron las soluciones dadas, con el fin de resolver las posibles dificultades y, a la misma vez, desarrollar el conocimiento del contenido de los participantes.

A continuación, se presenta los cuatro ítems que componían el cuestionario desarrollado para este estudio. La elaboración de los dos primeros ítems se ha realizado tomando como referencia los enunciados de los problemas de inferencia estadística que se han propuesto en las PAU. Se trata de problemas de respuesta abierta, propios de la investigación cualitativa, que permiten comprender con más profundidad los razonamientos y conocimientos de los estudiantes.

Ítem 1: Contraste de hipótesis

En primer lugar, se presentó un problema sobre contraste de hipótesis, cuyo enunciado es el que se presenta a continuación.

Ítem 1. En el punto 11 hemos visto que la esperanza media en el conjunto de países es 69,2 años y su desviación típica 10. Se ha tomado una muestra de 16 países europeos y se ha obtenido una esperanza de vida media de 78 años. Suponiendo que la variable esperanza de vida sigue la distribución normal, plantee un contraste de hipótesis, con un nivel de significación del 5%, para analizar si la esperanza media de vida en Europa es mayor que la esperanza media de vida en el conjunto de países. Explique a un estudiante con detalle los pasos que ha seguido en este contraste de hipótesis.

Con la información facilitada en el enunciado se pedía completar todos los pasos de un contraste de hipótesis unilateral de una cola para la media de una población cuya varianza es conocida, contenido incluido en el currículo anterior para Bachillerato (MEC, 2007). Pretendemos evaluar, mediante la respuesta a esta tarea, en qué medida los futuros profesores son capaces de cumplimentar los siguientes puntos:

1. *Planteamiento de las hipótesis estadísticas* con atención a dos puntos complementarios: a) Si las hipótesis se plantean en términos del parámetro, son complementarias y excluyentes, cubren el espacio paramétrico y se discrimina el papel de las hipótesis nula y alternativa, tratando de analizar la posible aparición de errores citados en Cañadas et al. (2012), Saldanha y Thompson (2007) o Vallecillos (1994); b) Si se indica que las hipótesis corresponden a un contraste unilateral y si el contraste se desarrolla consistentemente con el tipo de hipótesis planteadas, analizando posibles errores identificados por Vallecillos (1994).
2. *Elección consistente de un contraste de hipótesis* que corresponda a las hipótesis planteadas.
3. *Procedimiento general* que emplean los estudiantes y desarrollo del mismo, estudiando si se completan todos los pasos e identificando posibles errores en el procedimiento.
4. Cálculos realizados en la *determinación del p-valor* e interpretación del mismo, tratando de observar interpretaciones incorrectas descritas por Birnbaum (1982), Falk y Greenbaum (1995) o Krauss y Wassner (2002), entre otros.
5. *Interpretación y contextualización* de los resultados obtenidos, ya que la fase interpretativa se ha mostrado difícil en trabajos como el de Arteaga (2011).

Ítem 2: Construcción e interpretación de un intervalo de confianza

Puesto que la conclusión final del ítem 1 es que se debe rechazar la hipótesis planteada, surge la pregunta de interés, ¿en qué márgenes variará el valor verdadero del parámetro con una cierta confianza? Los problemas de intervalo de confianza se han propuesto casi sin excepción en todas las pruebas de acceso a la universidad y en las primeras ediciones de la nueva reválida de Bachillerato. Como se ha indicado, por ejemplo, en el Capítulo 2, los intervalos de confianza son recomendados ampliamente para complementar los contrastes de hipótesis (APA, 2010). De ello que sea necesario asegurar su comprensión por los futuros profesores. El segundo ítem tiene el siguiente enunciado:

Ítem 2. Determine un intervalo de confianza del 95% para la esperanza media de vida en los países europeos. ¿Cómo explicaría a un estudiante el significado de un intervalo de confianza y alguna de sus propiedades?

Respecto a este ítem el interés se centra en analizar los puntos siguientes:

1. *Planteamiento del intervalo.* Selección de un intervalo que corresponda al enunciado del problema, eligiendo tanto la distribución muestral como el estadístico muestral adecuado.
2. *Cálculo correcto del intervalo.* Observación de los pasos a seguir en la determinación de los extremos, analizando las posibles dificultades con la tipificación o lecturas de tablas descritas por Espinel et al. (2007).
3. *Interpretación* del significado del intervalo de confianza. Comprensión del carácter aleatorio de los extremos del intervalo de confianza evitando los errores descritos por Behar (2001), Cumming y Fidler (2005), Fidler y Cumming (2005) u Olivo (2008).
4. *Definición* por parte de los futuros profesores sobre el término en estudio. Análisis de las definiciones aportadas que permitirá estudiar con más detalle la comprensión del significado del intervalo de confianza.
5. *Propiedades fundamentales* que recuerdan los futuros profesores en relación a los intervalos de confianza.

Ítem 3 y 4: Análisis de los posibles errores que pueden cometer estudiantes de bachillerato

Puesto que el conocimiento del profesor ha de ir más allá del simple conocimiento matemático, es importante asegurar que adquiere algunos conocimientos didácticos. En particular, en este tema se centra la atención en analizar el conocimiento que poseen en relación a los posibles errores que podrían cometer los estudiantes de ese nivel educativo, es decir, se pretende evaluar una parte de la faceta cognitiva de su conocimiento didáctico. Aunque somos conscientes de que esta faceta incluye muchos más puntos que los considerados en el presente estudio, debido a la escasez de investigación sobre el tema, se aporta información original. Las preguntas planteadas

al respecto son las siguientes:

Ítem 3. ¿Podrías indicar algún error que podrían cometer los estudiantes al realizar un contraste de hipótesis?

Ítem 4. ¿Y al construir o interpretar un intervalo de confianza?

Esperamos que la experiencia de haber resuelto ellos mismos los dos primeros problemas les ayude a describir algunos posibles errores. Compararemos estos errores con los descritos en la investigación sobre el tema en el Capítulo 2.

4.4.3. Método de análisis de los datos

Cada uno de los participantes en el estudio completó por escrito el cuestionario, una vez que el profesor de la asignatura les explicó la finalidad del mismo y resolvió las dudas sobre la forma de completarlo. Los estudiantes mostraron su interés por el tema, completándolo antes de finalizar el tiempo disponible para ello. Una vez recogidos todos los cuestionarios se procedió a realizar un análisis de contenido de las respuestas a cada uno de los apartados.

Al igual que en el Estudio 1, se utiliza el análisis de contenido (Krippendorff, 2013) que permite realizar inferencias válidas a partir del análisis de textos. El análisis de contenido es una técnica que busca investigar sobre la naturaleza del discurso, es propio de la investigación cualitativa, permitiendo el estudio sistemático y la realización de inferencias sobre el contenido de documentos escritos (Neuendorf, 2016).

Se utiliza el estudio sistemático de documentos escritos, seleccionando del informe entregado por cada estudiante los apartados que correspondan a cada una de las variables que se analizan en el trabajo. Para realizar el análisis, se ha recurrido a nuestro conocimiento previo sobre el tema, apoyado en la revisión de antecedentes presentados en el Capítulo 2. Se siguieron las siguientes fases:

1. Para cada uno de los cuestionarios y cada ítem se analiza los tipos de respuestas diferenciados en varios apartados. Para ello se efectuaron varias lecturas cuidadosas para determinar la semejanza o diferencias de respuestas al mismo apartado.

2. Mediante un proceso cíclico e inductivo, se compararon las respuestas del mismo apartado hasta llegar a un sistema de categorías (unidad secundaria de análisis). El equipo de investigación revisó las categorías.
3. En cada una de las categorías de respuestas encontradas, se muestra un ejemplo realizado por uno de los participantes.

Finalmente se presenta la tabla de frecuencias de respuestas diferentes en el apartado y una síntesis de los resultados.

4.5. CONOCIMIENTO MATEMÁTICO SOBRE EL CONTRASTE DE HIPÓTESIS

Como se ha indicado, el conocimiento matemático sobre el contraste de hipótesis se evalúa a través de la resolución de un problema completo similar a los propuestos en las PAU. Para el análisis de las respuestas de los futuros profesores de Educación Secundaria Obligatoria y Bachillerato en este ítem la solución aportada se ha dividido en varios apartados referidos al planteamiento de las hipótesis elección del tipo de contraste, elección del tipo de contraste, desarrollo del procedimiento seguido, determinación del p -valor e interpretación de los resultados. A continuación, se analiza cada uno de estos apartados.

4.5.1. Planteamiento de las hipótesis

En primer lugar, se analizaron las hipótesis planteadas por los estudiantes que permitieron describir las diferentes categorías de respuestas encontradas, comenzando por las correctas, seguidas de las parcialmente correctas. Para cada categoría se presenta un ejemplo y se realiza el análisis semiótico del mismo. Una vez descritas las categorías se presentan y discuten los resultados en esta parte del problema. Para plantear correctamente las hipótesis se deben seguir los siguientes pasos:

1. Reconocer que se trata de un contraste sobre el parámetro media de una variable con distribución normal para una sola muestra, donde la desviación estándar es conocida.
2. Diferenciar entre el valor teórico de la media poblacional o parámetro ($\mu_0 = 69,2$) y el valor del estadístico media muestral observado ($\bar{x} = 78$).

3. Identificar la hipótesis nula y la alternativa respecto al parámetro poblacional presente en el problema, decidiendo que se trata de un contraste unilateral a derecha.
4. Plantear las hipótesis en forma adecuada utilizando la notación conveniente. Recordemos que, en un contraste de hipótesis, la hipótesis nula es planteada para ser rechazada ($H_0: \mu \leq 69,2$) y la hipótesis alternativa es la hipótesis contraria a la hipótesis nula ($H_1: \mu > 69,2$) (Batanero, 2000). En nuestro caso, se trata de realizar un contraste unilateral pues la región crítica es construida a un solo lado; es un contraste a derecha debido a que la hipótesis alternativa es que la media es superior a 69,2 y como consecuencia, la región crítica, o región de rechazo, se encuentra a la derecha.

Para la variable *Planteamiento de la hipótesis* se han definido 11 categorías: CH1, Planteamiento correcto; CH2, Planteamiento correcto de las hipótesis, pero sin utilizar la simbolización adecuada; CH3, Planteamiento correcto de las hipótesis con simbolización incorrecta; CH4, Las hipótesis no cubren el espacio paramétrico; CH5, Confusión entre ambas hipótesis; CH6, Confusión entre ambas hipótesis y entre la media muestral y poblacional; CH7, Planteamiento de una única hipótesis; CH8, Planteamiento de una única hipótesis y confusión entre la media muestral y poblacional; CH9, Planteamiento de las hipótesis en términos de proporciones; CH10, Inclusión del valor de la media poblacional en la hipótesis alternativa y CH11, Inclusión del valor de la media poblacional en ambas hipótesis. A continuación, se detallan cada una de ellas acompañadas de una respuesta dada por alguno de los sujetos participantes en el estudio.

CH1. *Planteamiento correcto*. Supone reconocer que el problema propuesto es un contraste sobre el valor medio de una variable en una población con distribución normal $N(\mu, \sigma)$, donde la desviación típica σ es conocida, como se especifica en el enunciado del problema. Se debe diferenciar en el enunciado el valor teórico de la media poblacional o parámetro ($\mu_0 = 69,2$) del valor del estadístico media muestral observada ($\bar{x} = 78$). A continuación, se ha de identificar la hipótesis nula y la alternativa correctas, teniendo en cuenta, que se trata de un contraste unilateral a derecha, ya que solo se está en los valores superiores a uno dado, es decir, en valores medios de la esperanza de vida al nacer superiores a 69,2.

Dado que la hipótesis nula de un contraste es la hipótesis que se desea rechazar (Batanero, 2000), en este caso la hipótesis nula establece que la esperanza media de vida en los países europeos es menor o igual que la esperanza de vida general en el global de países y, como consecuencia, la hipótesis alternativa será la contraria. Un ejemplo de planteamiento correcto es la respuesta del participante ACG, donde se observa que ambas hipótesis son complementarias y cubren el espacio paramétrico. Esta propiedad es importante pues permite finalizar el contraste con una decisión, ya que, si las hipótesis no fuesen complementarias, el resultado de la decisión podría ser ambiguo (Vallecillos, 1994).

$$H_0: \mu_e \leq 69,2; H_1: \mu_e > 69,2 \text{ (Respuesta de ACG).}$$

CH2. *Planteamiento correcto de las hipótesis, pero sin utilizar la simbolización adecuada.* Se han considerado en esta categoría aquellos sujetos que identifican correctamente que se trata de un contraste sobre la media de una población con distribución normal y que discriminan adecuadamente la hipótesis nula y alternativa, siendo ambas complementarias y planteadas en términos del parámetro, correspondientes a un contraste unilateral. Pero, en lugar de utilizar la notación adecuada, posiblemente por no recordarla, describen las hipótesis mediante uso de un lenguaje coloquial, sin simbolizarlas, como el caso que se describe a continuación. Observamos en este ejemplo una falta de correspondencia entre los objetos matemáticos y los símbolos que los representan, es decir, un conflicto semiótico representacional:

La hipótesis nula es que la esperanza de vida media en Europa es menor o igual que 69,2. Por tanto, la hipótesis alternativa es que la esperanza de vida media en Europa es mayor que 69,2 (Respuesta de AGG).

CH3. *Planteamiento correcto de las hipótesis con simbolización incorrecta.* Algunos futuros profesores llevan a cabo correctamente todos los pasos descritos en la categoría CH1. Es decir, plantean las hipótesis nula y alternativa adecuadas en términos del parámetro y establecen hipótesis complementarias. Sin embargo, hacen uso de una notación no empleada habitualmente, como ocurre en la respuesta dada por MTF en la que no queda claro si se diferencia correctamente la media muestral y poblacional existiendo una confusión implícita. Al igual que el caso anterior, el estudiante no es capaz de asociar una representación adecuada a los objetos matemáticos que utiliza y por tanto presenta un conflicto semiótico representacional. El uso de una simbolización

incorrecta para plantear las hipótesis también fue descrito en el estudio de Cañadas et al. (2012).

Llamamos M_e a la esperanza media de vida en Europa. $H_0: M_e \leq 69,2$ y $H_1: M_e > 69,2$ (Respuesta de MTF).

CH4. *Las hipótesis no cubren el espacio paramétrico.* El hecho de que el participante plantea hipótesis no complementarias implica que el proceso podría acabar en falta de decisión si se obtuviese un valor del estadístico en la región no cubierta por ambas hipótesis. En el siguiente ejemplo, se utiliza en el planteamiento de la hipótesis nula el signo de desigualdad “menor estricto”, sin considerar la igualdad de la media de la población al valor hipotético en la hipótesis nula. Este olvido es importante, pues la distribución muestral del estadístico se calcula precisamente en el supuesto de que la media de la población es igual al valor hipotético. El error de construir hipótesis no complementarias fue ya descrito por Cañadas et al. (2012) y Vallecillos (1999). En esta situación se detecta un conflicto semiótico de tipo conceptual al considerar hipótesis no complementarias y otro consistente en no comprender el papel que tiene el valor puntual de la hipótesis nula en la elección de la distribución muestral del estadístico, que depende de dicho valor puntual. A continuación, se incluye una respuesta de este tipo:

Planteamos el contraste de hipótesis donde se establece la hipótesis nula, H_0 (como comprobación por contraste) y la hipótesis alternativa, H_1 .
 $H_0: \mu_E < \mu \rightarrow \mu_E < 69,2$; $H_1: \mu_E > \mu \rightarrow \mu_E > 69,2$ (Respuesta de LUG).

CH5. *Confusión entre ambas hipótesis.* Cuando el estudiante identifica correctamente que el parámetro a contrastar es la media de la diferenciando el valor del parámetro (media poblacional) y del estadístico (media muestral) y también deduce, a partir del enunciado, que se trata de realizar un contraste unilateral, pero, sin embargo, intercambia las hipótesis nula y alternativa (véase la respuesta de AGB). Supone un nuevo conflicto semiótico conceptual consistente en confundir las dos hipótesis. Subyace una falta de comprensión del papel de cada una de las hipótesis en el contraste.

Dada una población sobre la que se observa la variable X , tal que X : esperanza de vida $\rightarrow N(69,2,10)$. Tenemos (X_1, \dots, X_{16}) una m.a.s. de $X \sim \bar{x} \rightarrow N(69,2,10/\sqrt{16})$ (distribución de la media muestral). Queremos contrastar a un nivel de significación $\alpha = 0,05$ la hipótesis nula frente a la alternativa. Enunciamos pues estas hipótesis. $H_0: \mu \geq 69,2$; $H_1: \mu < 69,2$ (Respuesta de AGB).

La confusión entre los papeles de las hipótesis nula y alternativa fue descrita por Cañadas et al. (2012) y Vallecillos (1999). Esta confusión puede ser consecuencia de una falta de comprensión de la lógica del contraste de hipótesis, lo que sería un conflicto semiótico de tipo procedimental. Esta lógica es más similar (aunque no idéntica) a la prueba deductiva por contradicción usual en matemáticas, pues se trata de rechazar una hipótesis. Sin embargo, en la mayor parte de las demostraciones que encuentran los estudiantes en otros temas matemáticos, la hipótesis se plantea para ser demostrada, no para ser rechazada, lo que puede justificar la confusión que tienen para el contraste de hipótesis. Además, mientras que en una prueba por contradicción un contraejemplo supone negar la hipótesis que se quiere rechazar, en inferencia un contraejemplo no informa siquiera sobre la probabilidad de que la hipótesis sea falsa, en la metodología frecuencial (Gigerenzer, 1993).

CH6. *Confusión entre ambas hipótesis y entre la media muestral y poblacional.* Al conflicto semiótico anterior, se añade en este caso la confusión entre media de la muestra y la media de la población, es decir, entre estadístico y parámetro, citada en Harradine et al. (2011), Schuyten (1991) y Vallecillos (1999). En esta situación, el estudiante usa el valor de la media muestral para plantear el contraste, como en el siguiente ejemplo, donde se indica explícitamente que la media de la muestra es 69,2 (siendo en realidad este valor el correspondiente a la media de la población) y plantea su hipótesis utilizando dicha media. Además del conflicto consistente en confundir las dos medias, hay otro por no comprender la lógica del contraste. No tiene sentido plantear una hipótesis sobre el valor muestral, puesto que no se trata de un valor hipotético, sino de una constante.

$\bar{x} = 69,2$ años, $\sigma = 10$, $n=16$, $\mu_0 = 78$, $\alpha = 0,05$. Tenemos el contraste $\begin{matrix} H_0: \mu \geq 78 \\ H_1: \mu < 78 \end{matrix}$ (Respuesta de DCC).

CH7. *Planteamiento de una única hipótesis.* Se consideran aquellos casos en los que el estudiante olvida plantear una de las dos hipótesis del contraste. En el siguiente ejemplo, sólo se plantea la hipótesis nula y, como consecuencia, se dificulta saber si la hipótesis alternativa está siendo considerada implícitamente. Es decir, no se comprende la necesidad de incluir la hipótesis alternativa en el planteamiento. Por tanto, se produce un conflicto procedimental. No obstante, explícitamente se reconoce el papel que juega la hipótesis nula.

A la hora de plantear la hipótesis de contraste, debe establecerse la hipótesis nula (H_0) como hipótesis de análisis, siendo ésta la que contradice lo que se quiere probar. Como en este supuesto se pretende probar que los países europeos tienen de media mayor esperanza de vida que el conjunto de la población, la hipótesis nula será la que rechace esta suposición, es decir, que la media de esperanza de vida en los países europeos NO es mayor que la media de esperanza de vida en el conjunto de la población. Como el valor de media que se tiene por cierto es el de la variable, la hipótesis H_0 será: $H_0 \leq 69,2$ (Respuesta de VRM).

CH8. *Planteamiento de una única hipótesis y confusión entre la media muestral y poblacional.* Se unen en esta respuesta dos de los conflictos semióticos anteriormente comentados. Por un lado, el contraste es planteado utilizando el valor de la media muestral como valor a contrastar, lo que supone confusión entre media muestral y poblacional, y además se plantea una única hipótesis de contraste. Ejemplo de esta categoría se observa en la siguiente respuesta, en la que no obstante se identifica correctamente la variable en estudio.

Nuestra variable será la esperanza de vida y nuestra hipótesis nula que la media es de 78 (Respuesta de DGM).

CH9. *Planteamiento de las hipótesis en términos de proporciones,* quedando confuso a qué proporción se refiere. Lo vemos en el ejemplo que sigue, donde se describen dos hipótesis utilizándose una simbología que hace referencia a la proporción y no a la media. La confusión implícita es considerar una variable aleatoria continua (la esperanza media de vida) como discreta. Este tipo de confusión no ha sido descrita en trabajos previos. Igualmente hay un conflicto semiótico al no diferenciar media y proporción.

Proponemos entonces dos posibles situaciones, una que verifique la condición y otra que no. Nuestra hipótesis nula será H_0 , y la contraria será H_1 . Estas hipótesis se verificarán si: $H_0: p_{EV} > p$; $H_1: p_{EV} < p$ (Respuesta de MMR).

CH10. *Inclusión del valor de la media poblacional en la hipótesis alternativa.* Se consideran en esta categoría aquellas respuestas en las que el valor hipotético del parámetro es utilizado para describir la hipótesis alternativa. Aunque la probabilidad de un valor puntual de una variable aleatoria continua es igual a cero, conceptualmente el valor supuesto de la media poblacional se debiera incluir en la hipótesis nula, porque es dicho valor el que se utiliza para definir la distribución muestral del estadístico. En consecuencia, el estudiante no comprende el papel que en esta distribución muestral tiene el valor supuesto puntual. Este conflicto semiótico conceptual no se ha descrito

en los trabajos previos analizados y se muestra en la respuesta que reproducimos:

Planteamos el contraste de hipótesis: $\begin{cases} H_0: \mu_e < 69,2 \\ H_1: \mu_e \geq 69,2 \end{cases}$ (Respuesta de CMM).

CH11. *Inclusión del valor de la media poblacional en ambas hipótesis.* El participante AGP no es consciente de que el valor hipotético de la media poblacional (en nuestro caso 69,2) debe ser considerado únicamente a la hora de definir la hipótesis nula. En esta situación observamos que el participante incluye dicho valor en ambas hipótesis obteniendo un contraste en el que no se verifica la propiedad de complementariedad entre ambas hipótesis ya que 69,2 corresponde a la intersección de las regiones definidas por H_0 y H_1 . Por consiguiente, por un lado, aparece el conflicto de suponer hipótesis no complementarias y por otro, de asignar el valor hipotético puntual a la hipótesis alternativa.

H_0 = hipótesis nula: $H_0 \equiv \mu_2 \leq 69,2$

H_1 = hipótesis alternativa: $H_1 \equiv \mu_2 \geq 69,2$ (Respuesta de AGP).

La Tabla 4.5.1 recoge los resultados referentes a este apartado, donde aproximadamente un 75% de los participantes plantean correctamente las hipótesis (CH1, CH2, CH3). Estos resultados mejoran los citados en investigaciones con otros tipos de estudiantes, lo que indica una mejor formación sobre el concepto de hipótesis en aquellos que se preparan como futuros profesores de Educación Secundaria Obligatoria y Bachillerato.

El resto de participantes muestran algunos de los errores descritos en investigaciones previas (Espinell et al., 2007, Harradine et al., 2011, Vallecillos, 1999), como, por ejemplo, formular hipótesis no excluyentes, que no cubren el espacio paramétrico o confundir la hipótesis nula o alternativa, así como el estadístico y el parámetro. También se han hallado errores nuevos, como expresar las hipótesis en términos de proporción (CH9) o incluir el valor hipotético en la hipótesis alternativa (CH10).

En la Tabla 4.5.2 se desglosan y se agrupan estos errores por los conflictos semióticos que subyacen en ellos que implican que algunos de los estudiantes que se preparan como futuros profesores los presentan. Los más frecuentes son los conflictos

semióticos de tipo conceptual, donde los estudiantes asignan a los conceptos asociados a las hipótesis estadísticas propiedades que no tienen o los confunden unos con otros.

Tabla 4.5.1.

Distribución de Respuestas en el Planteamiento de las Hipótesis

Código	Respuesta	Frecuencia	Porcentaje
CH1	Planteamiento correcto	49	67,2
CH2	Planteamiento correcto de las hipótesis, pero sin utilizar la simbolización adecuada	4	5,5
CH3	Planteamiento correcto de las hipótesis con simbolización incorrecta	2	2,7
CH4	Las hipótesis no cubren el espacio paramétrico	2	2,7
CH5	Confusión entre ambas hipótesis	3	4,1
CH6	Confusión entre ambas hipótesis y entre la media muestral y poblacional	3	4,1
CH7	Planteamiento de una única hipótesis	2	2,7
CH8	Planteamiento de una única hipótesis y confusión entre la media muestral y poblacional.	1	1,4
CH9	Planteamiento de las hipótesis en términos de proporciones	1	1,4
CH10	Inclusión del valor de la media poblacional en la hipótesis alternativa	4	5,5
CH11	Inclusión del valor de la media poblacional en ambas hipótesis	2	2,7
Total		73	100

Ello implica un conocimiento mecánico del planteamiento de las hipótesis sin una comprensión profunda del porqué de las características y forma de plantearlas. En todo caso son pequeños los porcentajes de los estudiantes que los comenten, aunque sería importante erradicarlos, pues se trata de futuros profesores.

4.5.2. Elección del tipo de contraste

Una vez planteadas las hipótesis, el paso siguiente sería elegir un contraste consistente con las hipótesis planteadas. Por ello se debe decidir si se trata de un contraste unilateral o bilateral, pues esta decisión condiciona el cálculo posterior del p -valor o de la región de rechazo e igualmente los cálculos en otros procedimientos empleados por los futuros profesores. En consecuencia, se analizó si se explicita el tipo de contraste que se va a utilizar y si dicho tipo es consistente con las hipótesis establecidas y el procedimiento utilizado. No se ha encontrado este tipo de análisis en investigaciones previas, por lo que este apartado constituye una aportación original.

Tabla 4.5.2.

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en la Formulación de las Hipótesis

Conflicto	Categorías	Frec.	Porcentaje
<i>Conceptual</i>			
1. Considerar hipótesis no complementarias en el contraste de hipótesis	CH4, CH11	4	5,4
2. No comprender el papel del valor puntual de la hipótesis nula en la elección de la distribución muestral del estadístico	CH4	2	2,7
3. Confundir la hipótesis nula y alternativa, y por tanto sus papeles en el contraste	CH5, CH8	4	5,4
4. Confundir estadístico (media muestral o proporción muestral) y parámetro (media poblacional o proporción poblacional)	CH5, CH6, CH8	7	9,6
5. Confundir variable aleatoria continua y discreta	CH9	1	1,4
6. No se comprende el papel del valor supuesto puntual en la hipótesis nula y por tanto en la definición de la distribución muestral del estadístico, asignando el valor puntual a la hipótesis alternativa	CH10, CH11	6	8,2
<i>Procedimental</i>			
1. Falta de comprensión o incorrecta aplicación de la lógica del contraste de hipótesis	CH5, CH6	6	8,2
2. Olvida plantear una de las dos hipótesis en un contraste	CH7, CH8	3	4,1
<i>Representacional</i>			
1. El estudiante no asocia los objetos matemáticos y los símbolos que los representan	CH2, CH3	6	8,2

En la respuesta correcta se espera que el estudiante especifique que se trata de un contraste sobre la media de una población normal. Además, debería identificar el contraste como unilateral derecho, porque la región crítica debe situarse a la derecha del valor hipotético (intervalos de valores superiores a 69,2). Se han definido las siguientes seis categorías en el análisis de las respuestas a este apartado.

CP1. *Especificar que se trata de un contraste unilateral derecho.* Junto a la descripción consistente de las hipótesis y la elección correcta del procedimiento a seguir, el estudiante sitúa la región crítica a la derecha del valor hipotético. Ejemplo de esta categoría es la respuesta del participante FLS, quien pone de manifiesto su conocimiento de la terminología asociada, la relación de las hipótesis planteadas con el tipo de contraste y con el cálculo posterior del p -valor. Igualmente realiza una definición correcta del significado del p -valor.

Por tanto, nuestro contraste podemos plantearlo como: $H_0: \mu_e \leq 69.2$; $H_1: \mu_e > 69.2$. Como nuestro test es unilateral a la derecha calculamos el p-valor al valor observado, es decir, la probabilidad de obtener el resultado que hemos observado si H_0 fuese cierta. $P[Z \geq Z_{78}] = P[Z \geq 3,52] = 1 - P[Z \leq 3,52] = 0.00022 < \alpha = 0.05$ (Respuesta de FLS).

CP2. *Indicar que se trata de un contraste unilateral izquierdo.* El contraste de hipótesis se denomina unilateral izquierdo cuando la región crítica o de rechazo se encuentra a la izquierda o por debajo del valor hipotético, es decir, se refiere a valores menores que el supuesto. Algunos estudiantes confunden el tipo de contraste a emplear, eligiendo un contraste unilateral izquierdo, lo que implica un conflicto semiótico conceptual. Un ejemplo es la respuesta dada por ALM, quien, además de establecer las hipótesis en función del valor de la media muestral, intercambia las hipótesis considerando como $H_1: \mu < 78$. Por tanto, muestra adicionalmente otros dos conflictos conceptuales ya descritos. El estudiante tipifica para encontrar el valor crítico en la distribución normal y define la región crítica a la izquierda del valor hipotético, en vez de a la derecha como $(-\infty, -1,645)$, planteando un contraste unilateral izquierdo, que explicita en su descripción.

Se trata de un contraste de hipótesis unilateral por la izquierda: $H_0: \mu \geq 78$; $H_1: \mu < 78$. A continuación, calculamos los puntos críticos que nos darán las regiones de aceptación y rechazo. El nivel de significación es del 5%; por tanto $\alpha=0,05$; luego $1-\alpha=0,95$. Mirando la tabla de la $N(0,1)$ vemos que esta probabilidad no viene. Los valores más próximos son 0,9495 y 0,9505 que corresponden a 1,64 y 1,65. La región crítica o de rechazo sería $(-\infty, -1,645)$ (Respuesta de ALM).

CP3. *Seleccionar un contraste unilateral sin indicar la dirección de la región crítica, y aplicar un contraste unilateral derecho.* Como ejemplo, se muestra la respuesta dada por CGG en la que se emplea notación incorrecta (trata de calcular la probabilidad de un punto aislado de la distribución); a pesar de ello se calcula correctamente el p-valor correspondiente a un contraste unilateral derecho, es decir, la probabilidad de obtener valores mayores o iguales que el observado 78. En consecuencia, aunque el estudiante ha elegido correctamente el tipo de contraste, consistente con las hipótesis planteadas, no recuerda la terminología, lo que supone un conflicto semiótico representacional.

Se trata de un contraste unilateral, pues la región crítica se construirá a un único lado. [...]. Estudiamos la probabilidad de que la media muestral sea 78 dada la hipótesis nula calculando el caso más desfavorable: $P(\text{media muestral}=78 | \mu=69,2)$. Tipificamos para tener una variable que siga una distribución $N(0,1)$. $Z = \frac{78-69,2}{10/4} = 3,52$. Observando la tabla de la distribución normal estándar, tenemos 0,9998 para $Z = 3,52$. Así, la probabilidad de obtener Z es de $1 - 0,9998 = 0,0002$ (Respuesta de CGG).

CP4. *No especificar el tipo de contraste, pero desarrolla un contraste unilateral derecho.* Es la respuesta más frecuente, donde el estudiante no siente la necesidad de indicar el tipo de contraste que está realizando. En el siguiente ejemplo, se observa que, aunque no define el p -valor como una probabilidad condicionada, el cálculo de este viene asociado a un contraste unilateral derecho, esto es, los valores del estadístico mayores que el valor observado. Asimismo, en el ejemplo mostrado se observa un conflicto consistente en definir la distribución muestral como una distribución no condicionada por el valor hipotético:

El contraste de hipótesis a realizar sería el siguiente: $H_0: \mu \leq 69,2$; $H_1: \mu > 69,2$. Para resolverlo, calculamos el p -valor, que es la probabilidad de que la media muestral sea mayor a 78 años de esperanza de vida: $p_{0,05} = P(\bar{X} > 78) = P\left(Z > \frac{78-69,2}{10/\sqrt{16}}\right) = P(Z > 3,52) = 1 - P(Z < 3,52) = 1 - 0,99978 = 0,00022$. Como esta probabilidad es muy pequeña ($p_{0,05} < 0,05$), rechazamos la hipótesis nula, y podemos suponer que la media de los países europeos es mayor a la del conjunto de países (Respuesta de MSM).

CP5. *No especificar el tipo de contraste y desarrollar un contraste bilateral.* En este caso la respuesta no es consistente con las hipótesis planteadas, pues los cuantiles $-z_{\alpha/2}$ y $z_{\alpha/2}$ corresponden a los valores que determinan la región de rechazo y aceptación de un contraste bilateral. Se observa un conflicto conceptual consistente en confundir un contraste unilateral con otro bilateral y otro que implica confundir la región crítica unilateral y bilateral.

Tomamos como H_0 la hipótesis nula como lo que quiero rechazar, en nuestro caso que la esperanza de vida media sea menor o igual que el conjunto de países. $H_0: \mu \leq 69,2$; $H_1: \mu > 69,2$. [...] $Z = \frac{(X_m - \mu) \times \sqrt{n}}{\sigma}$ se distribuye según $N(0,1)$. A nivel $\alpha = 0,05$ si tomamos la representación gráfica de $N(0,1)$ el punto a partir del cual se rechaza H_0 sería si $(78 - 69,2)/2,5$ es mayor que $z_{\alpha/2} = z_{0,025}$ o si es menor que $-z_{\alpha/2} = -z_{0,025}$ [...] (Respuesta de ECL).

CP6. *No especificar el tipo de contraste y no desarrollar los pasos.* Se consideran aquellas respuestas en las que no queda claro si el participante sabe realizar los pasos necesarios para la elaboración del contraste. A modo de ejemplo se muestra la respuesta dada por DGM donde no se detalla el tipo de contraste e indica únicamente que debe ser rechazada la hipótesis, la cual describe erróneamente ya que el valor de la media de la muestra no está dentro de un intervalo que ni siquiera construye.

Nuestra variable será la esperanza de vida y nuestra hipótesis nula que la media es de 78, como no nos sale dentro del intervalo, la rechazamos, por lo que estimamos que la esperanza de vida en Europa es mayor que en el conjunto de países que hemos escogido

(Respuesta de DGM).

En la Tabla 4.5.3 se presentan los resultados obtenidos en este apartado, donde se observa que la mayoría de los futuros profesores (54,8%) desarrolla un contraste unilateral derecho, consistente con las hipótesis establecidas, pero no indica de qué tipo de contraste se trata. Añadiendo los que desarrollan el contraste unilateral derecho, aunque no lo especifiquen, el 83% elige un procedimiento correcto. Son pocos los futuros profesores que confunden el contraste unilateral derecho e izquierdo. Por último, se ha observado que tres participantes no especifican y no llegan a desarrollar el contraste.

Tabla 4.5.3.

Distribución de Respuestas en el Tipo de Contraste Elegido

Código	Respuesta	Frecuencia	Porcentaje
CP1	Especificar que se trata de un contraste unilateral derecho.	3	4,1
CP2	Indicar que se trata de un contraste unilateral izquierdo	1	1,4
CP3	Seleccionar un contraste unilateral sin indicar la dirección de la región crítica y aplicar un contraste unilateral derecho	18	24,7
CP4	No especificar el tipo de contraste, pero desarrollar un contraste unilateral derecho	40	54,8
CP5	No especificar el tipo de contraste y desarrollar un contraste bilateral	8	11,0
CP6	No especificar el tipo de contraste y no desarrollar los pasos	3	4,1
Total		73	100

En la Tabla 4.5.4 se desglosa y agrupa las respuestas por los conflictos semióticos que subyacen en ellos que implican que algunos de los estudiantes que se preparan como futuros profesores los presentan.

Frecuentemente los participantes no usan la probabilidad condicional para definir la distribución muestral asociada al estadístico, lo que implica o bien que no se es consciente de la naturaleza condicional de dicha distribución o de una posible confusión entre probabilidad simple y condicional, conflicto detectado por Falk (1986). Por otro lado, la mayoría de estudiantes utiliza simbología incorrecta. El resto de conflictos tiene poca frecuencia.

Tabla 4.5.4.

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en la Elección del Tipo de Contraste

Conflicto	Categorías	Frec.	Porcentaje
<i>Conceptual</i>			
1. Confundir la hipótesis nula y alternativa, y por tanto sus papeles en el contraste	CP2	1	1,4
2. Confundir estadístico (media muestral o proporción muestra) y parámetro (media poblacional o proporción poblacional)	CP2	1	1,4
3. Confundir un contraste de hipótesis unilateral derecho con otro unilateral izquierdo	CP2	1	1,4
4. Confundir un contraste de hipótesis unilateral con otro bilateral	CP5	8	11,0
5. Confundir región crítica de un contraste de hipótesis unilateral y bilateral	CP5	8	11,0
6. No definir la distribución muestral como una distribución condicionada por el valor hipotético	CP4	40	54,8
<i>Representacional</i>			
1. No asociar los objetos matemáticos y su terminología	CP3, CP4	58	79,5

4.5.3. Desarrollo del procedimiento

Son varias las metodologías propuestas en inferencia estadística para contrastar una hipótesis, cada una de las cuáles condiciona el desarrollo del procedimiento a seguir (Batanero, 2000, Gigerenzer, 1993; Rivadulla, 1991). Puesto que los participantes provienen de diferentes centros universitarios, podrían haber estudiado diferentes metodologías de contraste de hipótesis, lo que nos llevó a interesarnos por el procedimiento empleado.

El estudio del método elegido por los participantes ha permitido observar que la mayoría de los futuros profesores (70%) siguió la metodología propuesta por Fisher (1935) basada en el cálculo de la probabilidad de obtener el valor observado del estadístico u otro más extremo (p -valor) y en el rechazo de la hipótesis nula si dicha probabilidad es inferior a un valor prefijado (nivel de significación), al suponer que en esta situación los datos aportan evidencia en contra de la hipótesis nula (Batanero, 2000; Inzunza y Jiménez, 2013).

En realidad, el método seguido caería dentro de lo que se ha denominado

metodología híbrida del contraste de hipótesis (Gigerenzer, 1993), puesto que se fija una hipótesis alternativa concreta, lo que no era recomendado por Fisher, pero es habitual en el método de Neyman y Pearson. Además, se fija de antemano el nivel de significación. No obstante, si los pasos en el desarrollo del procedimiento son correctos, se acepta esta forma de proceder como correcta, pues es la que habitualmente se enseña, a pesar de los problemas filosóficos asociados a la misma.

El resto de participantes siguió otros procedimientos (analizados en Rivadulla, 1991): usar el procedimiento correcto de Neyman-Pearson calculando las regiones críticas y de aceptación (16,4%), traducir el problema al cálculo de un intervalo de confianza (6,8%), emplear el método de máxima verosimilitud comparando la verosimilitud de los datos en los supuestos de ser cierta la hipótesis nula o la alternativa (4,1%) o no llegar a resolverlo (2,7%). Teniendo en cuenta la metodología elegida, a continuación, definimos una siete categorías en función al desarrollo consiguiente del procedimiento:

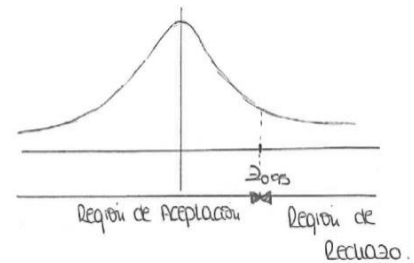
CD1. *Realizar un procedimiento correctamente desarrollado.* Se considera en esta categoría el desarrollo correcto del procedimiento, sea cual sea la metodología aplicada. La aplicación correcta de la metodología de Fisher supone definir el p -valor como la probabilidad de obtener un valor de la media muestral superior o igual a 78, bajo la suposición de que la hipótesis nula H_0 es cierta, es decir, asumiendo que la media poblacional toma el valor 69,2. Además, los estudiantes en esta categoría dedujeron correctamente la distribución muestral, tipificando el valor obtenido del estadístico para utilizar la tabla de la distribución Normal estándar. Este último paso no sería necesario si se usa la calculadora, pero los estudiantes siguen el procedimiento aprendido, basado en el uso de las tablas. Así, en el ejemplo que sigue, se muestra el especial cuidado en describir los pasos llevados a cabo.

$$P(\bar{X} \geq 78 | \mu = 69,2) = (\text{Tipifico}) = P\left(\frac{\bar{X}-69,2}{2,5} \geq \frac{78-69,2}{2,5}\right) = P(Z \geq 3,52) = 1 - P(Z \leq 3,52) = \\ (\text{Usando la tabla de la } N(0,1)) = 1 - 0,99978 = 0,00022 \text{ (Respuesta de EHB).}$$

También se incluye en esta categoría aquellos estudiantes que desarrollan correctamente el procedimiento en la metodología de Neyman y Pearson. Ejemplo de una aplicación correcta del procedimiento de Neyman y Pearson es la respuesta dada por el participante NSF, llevando a cabo el cálculo correcto del estadístico de contraste

$\left(Z = \frac{\bar{x}-69,2}{2,5} = \frac{78-69,2}{2,5} = 3,52 \right)$ y de la región de rechazo $(1,65, +\infty)$.

Se trata de un contraste unilateral por la derecha, con lo cual la región crítica está a la derecha del punto crítico $z_{0,95} = 1,65$. En este caso, el estadístico de contraste es $Z = \frac{\bar{x}-69,2}{2,5}$ que sigue una distribución $N(0,1)$. Por lo tanto, el valor observado del estadístico a partir de nuestra muestra es $Z_0 = \frac{78-69,2}{2,5} = 3,52$. Como $Z_0 > z_{0,95} = 1,65$ entonces rechazamos la hipótesis nula para un nivel de significación $\alpha = 0,05$. (Respuesta de NSF)



CD2. *Realizar un procedimiento correcto para un contraste bilateral.* Realizar un contraste bilateral, incluso correctamente, implica un error en el cálculo del p -valor, la región de rechazo, o función de verosimilitud, dependiendo del procedimiento seleccionado. La confusión entre el contraste unilateral y bilateral fue descrita por Espinel et al. (2007) y en este trabajo se ha interpretado como conflicto semiótico conceptual. Además, en este apartado, supone un conflicto semiótico procedimental, pues hay una confusión en el cálculo de la región crítica, o el intervalo de confianza, si se desarrolla este método. Como ejemplo, se presenta una respuesta donde se desarrolla el contraste mediante un intervalo de confianza del 95%, que viene asociado a un contraste bilateral, en vez de intervalo unilateral como sería lo correcto. En esta situación, en vez de contrastar la hipótesis de si la media de la población es menor o igual a 69,2 se contrasta la hipótesis de que la media de la población tome el valor 69,2 frente a que no tome dicho valor:

Planteo intervalo de confianza para $\alpha = 0,05$; $P(-z^* \leq Z \leq z^*) = 0,95$; $P\left(\bar{x} - 1,96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \times \frac{\sigma}{\sqrt{n}}\right) = 0,95$. [...] Una vez planteado el intervalo de confianza sustituyo los datos: [76,78; 79,23] (Respuesta de JGM).

CD3. *Realizar un procedimiento correcto usando la distribución t de Student en lugar de la distribución Normal.* Puesto que la distribución de partida es normal, de desviación típica conocida, la distribución muestral de la media es también normal. El uso de la distribución t de Student tanto en la realización de un contraste de hipótesis como en la determinación de los extremos de un intervalo de confianza para la media (o diferencia de medias) se realiza bajo la suposición de que la varianza poblacional es desconocida. Supone un conflicto semiótico conceptual al no recordar las condiciones de aplicación de uno u otro modelo probabilístico. La utilización de dicha distribución

carece de sentido en aquellos casos en los que la varianza de la población es conocida puesto que la distribución de la variable en estudio queda totalmente determinada. En la respuesta dada por ILB se observa que el uso de dicho modelo probabilístico implica la obtención de un valor crítico incorrecto (1,753), lo que supone un conflicto semiótico procedimental. Además, realiza la comparación entre dos valores provenientes de distintas distribuciones ($Z = 3,52$ es válido cuando el estadístico de contraste se distribuye normalmente mientras que $t_{\alpha} = 1,753$ es obtenido a partir de una distribución t de Student con n grados de libertad). Otro conflicto semiótico procedimental es comparar valores críticos obtenidos de diferentes distribuciones.

Mirando la tabla de la t de Student $t_{\alpha} = 1,753$. Como $Z = 3,52$, mirando el gráfico ($1,753 < 3,52$) cae en la zona de rechazo. Así que la hipótesis nula no es válida (Respuesta de ILB).

CD4. *Expresar el p -valor como probabilidad simple, aunque es calculado como probabilidad condicional.* Algunos futuros profesores definen el p -valor como una probabilidad simple, omitiendo la condición de que la hipótesis nula sea cierta en la simbolización utilizada. Sin embargo, al tipificar utilizan la distribución muestral correcta, que se deduce con la condición de ser cierta la hipótesis nula, y calculan correctamente su valor, como ocurre en el siguiente ejemplo. Aunque este error no afecta a los cálculos, supone una falta de comprensión del concepto de distribución muestral, que está condicionada por el valor hipotético, o al menos una imprecisión en la notación. Por tanto, el p -valor implicaría la existencia de un conflicto semiótico conceptual además de un conflicto semiótico representacional, que, a su vez, podría encubrir otro conflicto semiótico conceptual consistente en no definir la distribución muestral de la media en forma condicional.

$$P(\bar{X} \geq 78) = P\left(\frac{\bar{X}-69,2}{2,5} \geq \frac{78-69,2}{2,5}\right) = P\left(Z \geq \frac{8,8}{2,5}\right) = P(Z \geq 3,5). \text{ Y esto es lo mismo que } P(Z \geq 3,5) = 1 - P(Z < 3,5) \text{ (Respuesta de AMC).}$$

CD5. *No considerar explícitamente el valor observado del estadístico en el cálculo del p -valor.* Otros participantes en el estudio definen el p -valor como la probabilidad de que el estadístico tome valores superiores al observado, sin considerar dicho valor explícitamente. Este error tampoco afecta a los cálculos, pues la probabilidad de un valor aislado, en el caso de variables aleatorias continuas, es igual a cero, pero al igual que en la categoría anterior, supone una falta de comprensión del concepto de p -valor, definido como la probabilidad de un valor igual o más extremo que el observado. Un

ejemplo se expone a continuación:

Para resolver el contraste, calculamos la probabilidad de que la media muestral obtenida sea mayor de 78 años de esperanza de vida, ya que si se obtiene una probabilidad muy baja de que ocurra, se podrá rechazar la hipótesis nula. Lo vemos: $P(\bar{x} > 78) = P\left(\frac{\bar{x}-69,2}{2,5} > \frac{78-69,2}{2,5}\right) = P\left(Z > \frac{8,8}{2,5}\right) = P(Z > 3,5) = 0,00022$ (Respuesta de IPM).

CD6. *Definir el p-valor como la probabilidad de que el estadístico tome un valor puntual.* En esta categoría se han incluido las respuestas en las que se describe el p-valor como la probabilidad de que el estadístico tome el valor 78, en la hipótesis establecida, es decir, $P(\bar{X} = 78|\mu = 69,2)$. Asimismo, se ha considerado los casos en los que se indica la probabilidad simple, $P(\bar{X} = 78)$. Como en los casos anteriores, no se comprende la definición del p-valor; además, supone falta de comprensión de las propiedades de una variable continua. Subyacen dos conflictos semióticos, el primero definir el p-valor como una probabilidad puntual y el segundo la no comprensión de la variable aleatoria continua. Recordemos que, si X es una variable aleatoria de tipo continuo, la probabilidad de que X tome un valor concreto es cero, es decir, $P[X = x_0] = 0, \forall x_0 \in \mathbb{R}$. Así pues, si se considera el p-valor como $P(\bar{X} = 78|\mu = 96,2)$ o como $P(\bar{X} = 78)$, no se necesita uso ni del proceso de tipificación ni de la tabla de probabilidad, ya que dicha probabilidad sería nula. Este error se presenta en la respuesta del estudiante GPF:

Estudiamos la probabilidad de que media muestral sea 78 dada la hipótesis nula calculando el caso más desfavorable: $P(\text{media muestral} = 78|\mu = 69,2)$. Tipificamos para tener una variable que siga una distribución $N(0,1)$. $Z=(78-69,2)/(10/4)=3,52$. Observando la tabla de la distribución normal estándar, tenemos 0,9998 para $Z=3,52$. Así, la probabilidad de obtener Z es de $1-0,9998=0,0002$. (Repuesta de GPF).

CD7. *Intercambiar las regiones de aceptación y crítica al utilizar el método de máxima verosimilitud.* El error de intercambiar las regiones de aceptación y crítica de un contraste cuando se trabaja el método de Neyman-Pearson fue descrito por Vallecillos (1994), pero no lo hemos encontrado identificado para el método de máxima verosimilitud. En dicho método, para realizar el contraste de hipótesis se comparan las funciones de verosimilitud de los datos observados en la condición de ser cierta la hipótesis nula y de ser cierta la alternativa, aceptando la hipótesis que dé mayor verosimilitud a los datos. Algunos estudiantes resuelven el problema mediante este método, pero en algunos casos interpretan incorrectamente los resultados, por lo que llegan a una conclusión incorrecta, como ocurre en la respuesta dada por JSF, donde a

pesar de cumplirse la condición que apoya a la hipótesis nula se rechaza. Por tanto, subyace un conflicto semiótico conceptual consistente en confundir las regiones críticas y de aceptación y otro procedimental porque se aplica erróneamente el resultado de la razón de verosimilitudes:

Estamos ante un contraste sobre la media de una normal con varianza conocida. Por tanto, el test de razón de verosimilitud de tamaño $\alpha = 0,05$ es:

$$\varphi(X_1, \dots, X_{16}) = \begin{cases} 1 & \text{si } \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} < z_{1-\alpha} \\ 0 & \text{si } \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \geq z_{1-\alpha} \end{cases}$$

donde $z_{1-\alpha}$ es el valor de una distribución normal estándar de parámetros 0 y 1 que deja a su derecha una probabilidad de $\alpha/2$. Ahora, tenemos que $\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} = \frac{78-69,2}{10/\sqrt{16}} = \frac{8,8}{2,5} = 3,52$. Y además que $z_{1-\alpha} = z_{1-0,05} = z_{0,95} = 1,65$. Por tanto, $3,52 > 1,65$, luego $\varphi(X_1, \dots, X_{16}) = 0$, esto es, los datos no dan evidencia para rechazar la hipótesis nula. (Respuesta de JSF).

La Tabla 4.5.5 recoge las frecuencias y porcentajes de las categorías resultantes en el procedimiento desarrollado para resolver el contraste de hipótesis, donde resalta un porcentaje pequeño de procedimientos desarrollados con corrección total (15,1%), lo que se agrava teniendo en cuenta que se trata de futuros profesores. La mayoría (70%) basa su procedimiento en el cálculo del p -valor, y se ha observado un gran porcentaje de errores sobre este concepto (61,6%), el cual es interpretado como probabilidad simple por un 11% de participantes, como probabilidad puntual por un 39,7% y un 11% no incluyen el valor hipotético en el cálculo. Estos errores de interpretación del p -valor no han sido descritos en la investigación previa.

Además, se ha encontrado que un 17,8% han confundido el contraste unilateral y el contraste bilateral, error descrito previamente por Espinel et al. (2007) y Vallecillos (1994).

En la Tabla 4.5.6 se resumen los conflictos semióticos que subyacen en las anteriores respuestas, algunos de los cuales tienen porcentajes importantes, tratándose de futuros profesores. Por ejemplo, confundir un contraste unilateral y bilateral, definir p -valor como la probabilidad puntual y no comprender que la probabilidad puntual en una variable aleatoria continua es siempre igual a cero. Igualmente tienen porcentajes importantes el error en el cálculo de la región crítica y del intervalo de confianza.

Tabla 4.5.5.*Distribución de Respuestas en el Desarrollo del Procedimiento*

Código	Desarrollo del procedimiento	Frecuencia	Porcentaje
CD1	Realizar un procedimiento correctamente desarrollado	11	15,1
CD2	Realizar un procedimiento correcto para un contraste bilateral	13	17,8
CD3	Realizar un procedimiento correcto usando la distribución t de Student en lugar de la distribución Normal	1	1,3
CD4	Expresar el p -valor como probabilidad simple, aunque es calculado como probabilidad condicional	8	11,0
CD5	No considerar explícitamente el valor observado del estadístico en el cálculo del p -valor	8	11,0
CD6	Definir el p -valor como la probabilidad de que el estadístico tome un valor puntual	29	39,7
CD7	Intercambiar las regiones de aceptación y crítica al utilizar el método de máxima verosimilitud	3	4,1
Total		73	100,0

Tabla 4.5.6.*Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en el Desarrollo del Procedimiento*

Conflicto	Categorías	Frecuencia	Porcentaje
<i>Conceptual</i>			
1. Confundir un contraste de hipótesis unilateral con otro bilateral	CD2	13	17,8
2. Confundir región crítica y de aceptación	CD7	3	4,1
3. Confundir las condiciones en que se puede aplicar una distribución normal para la distribución de la media muestral	CD3	1	1,3
4. No definir la distribución muestral como una distribución condicionada por el valor hipotético	CD4	8	11,0
5. Definir el p -valor, sin incluir en su cálculo el valor hipotético	CD5	8	11,0
6. Definir el p -valor como probabilidad puntual	CD6	29	39,7
7. No comprender que la probabilidad puntual en una v.a. continua es siempre igual a cero	CD6	29	39,7
<i>Procedimental</i>			
1. Cálculo erróneo de la región crítica	CD2	13	17,8
2. Cálculo erróneo de intervalo de confianza	CD2	13	17,8
3. Cálculo incorrecto de valor crítico	CD3	1	1,3
4. Comparar valores críticos procedentes de diferentes distribuciones	CD3	1	1,3
5. Aplicar erróneamente el resultado de la razón de verosimilitudes	CD7	3	4,1
<i>Representacional</i>			
1. No asocia los objetos matemáticos y su terminología	CD4	8	11,0

4.5.4. Determinación del p -valor

Puesto que un alto número de participantes (51) recurren el cálculo del p -valor, y dadas las dificultades encontradas en los futuros profesores a la hora de definirlo correctamente, se completa el análisis con el estudio de la forma en que estos futuros profesores finalmente obtienen el p -valor, independientemente de cómo lo han definido, obteniendo las siguientes cuatro categorías:

CC1. *Calcular el p -valor mediante tipificación y usar la tabla de la normal estandarizada.* Estos futuros profesores definen correctamente el p -valor y a continuación tipifican el valor obtenido del estadístico para calcularlo. Por ello mostrarían una correcta comprensión de la tipificación y competencia en la lectura de la tabla de la distribución normal $N(0,1)$, como ocurre en la siguiente respuesta:

El estadístico de contraste $Z = \frac{78-69,2}{2,5} = 3,5$ sigue una distribución normal $N(0,1)$. [...]. Suponiendo cierta la hipótesis nula, la probabilidad de obtener $Z = 3,5$ es $p = 1 - 0,99865 = 0,00135$ (Respuesta de SS).

CC2. *Aproximar el p -valor por 0.* Teniendo dificultad en la lectura de las tablas de la distribución normal $N(0,1)$, algunos futuros profesores simplemente aproximan por cero el p -valor, lo que supone una falta de precisión, además de un conflicto procedimental por no ser capaz de utilizar las tablas estadísticas. Lo observamos en la respuesta dada por MAM que, además, define el p -valor como probabilidad puntual y mezcla probabilidades simples y condicionales, mostrando por tanto dos conflictos, el primero conceptual y el segundo procedimental:

$$P(\bar{X} = 78 | \mu = 69,2) = P\left(\frac{78-69,2}{2,5} \mid \mu = 69,2\right) = P(Z \geq 3,5) = 0 < 0,05 \text{ (Respuesta de MAM).}$$

CC3. *Aproximar el p -valor utilizando el porcentaje de casos incluidos en los intervalos centrales en una distribución normal.* Algunos futuros profesores tratan de paliar el olvido de la lectura de las tablas recordando las propiedades de simetría de la distribución normal respecto al valor medio μ , y el porcentaje de casos incluidos en los intervalos centrales de la distribución. Dicha propiedad consiste en que los intervalos $[\mu - \sigma, \mu + \sigma]$, $[\mu - 2\sigma, \mu + 2\sigma]$ y $[\mu - 3\sigma, \mu + 3\sigma]$ contienen aproximadamente el 68,2%, 95,4% y 99,6% de la distribución, respectivamente. En el siguiente ejemplo, aunque el estudiante no concluye correctamente el valor de la probabilidad que se está calculando, el participante hace uso de la mencionada propiedad. Por tanto, subyace un conflicto procedimental consistente en olvido de la lectura de las tablas de la distribución normal.

Además, expresa la distribución muestral en forma no condicionada, que es un conflicto conceptual:

Si tipificamos se tiene que $P(\bar{x} > 78) = P\left(\frac{\bar{x}-69,2}{2,5} > \frac{78-69,2}{2,5}\right) = P(Z > 3,5)$. Al tipificar trabajamos con $N(0,1)$. Se sabe que en la $N(0,1)$ se tiene que: $P(-1 < Z < 1) = 0,68$, $P(-2 < Z < 2) = 0,95$, $P(-3 < Z < 3) = 0,99$. Por tanto, teniendo en cuenta que se fija un nivel de significación del 5%, se obtiene que $P(Z > 3,5) = 0,005$ (Respuesta de CBF).

CC4. *Describir el cálculo del p-valor sin llegar a completarlo.* Algunos futuros profesores comienzan el procedimiento del contraste, pero no calculan el *p*-valor, sino que describen la forma de calcularlo, como se ve en la respuesta de AGP. Implícitamente subyace el olvido del cálculo de probabilidades en intervalos y la lectura de la tabla. Además, expresa la probabilidad como simple y no como condicional:

¿Cuál es la $P(X = 78 \mid \mu_2 = 69,2)$? Tipificamos $P = \frac{78-69,2}{2,5}$; $\mu_2 = 69,2$; $P(Z \geq 3,5)$ (Respuesta de AGP).

El análisis de la forma en que los 51 estudiantes basaron el cálculo del *p*-valor, ha permitido ver que el 70,6% de los mismos tipifican y realizan un uso correcto de la tabla de la distribución normal estándar (véase Tabla 4.5.7). El resto presenta ciertas dificultades aproximando a cero la probabilidad (7,8%), recurriendo a la propiedad de los intervalos centrales de una distribución normal (17,6%) y dos estudiantes no fueron capaces de llegar a obtenerlo. Estas dificultades de cálculo se unen a los errores de definición descritos en el apartado anterior y no identificados en la investigación previa.

Tabla 4.5.7.

Distribución de Respuestas en la Obtención del p-valor

Códigos	Método de obtención	Frecuencia	Porcentaje
CC1	Calcular el <i>p</i> -valor mediante tipificación y usar la tabla de la normal estandarizada	36	70,6
CC2	Aproximar el <i>p</i> -valor por 0	4	7,8
CC3	Aproximar el <i>p</i> -valor utilizando el porcentaje de casos incluidos en los intervalos centrales en una distribución normal	9	17,7
CC4	Describir el cálculo del <i>p</i> -valor sin llegar a completarlo	2	3,9
Total		51	100,0

Por el contrario, en el presente estudio no se ha hallado el error relacionado con el intercambio de los términos de la probabilidad condicional en la interpretación del *p*-

valor encontrado en investigaciones como las de Birnbaum (1982), Falk (1986) o Vallecillos (1994).

En la Tabla 4.5.8 se desglosan y agrupan las respuestas en el cálculo del p -valor por los conflictos semióticos identificados. El principal conflicto conceptual es no tener en cuenta la naturaleza condicional de la distribución muestral y un pequeño número de estudiantes atribuye una probabilidad puntual al p -valor. Es mayor la frecuencia de conflictos procedimentales, en particular la lectura correcta de las tablas de la distribución normal, también encontrado en la investigación de Espinel et al. (2007) en estudiantes de Bachillerato.

Tabla 4.5.8.

Frecuencia y Porcentaje de Estudiantes con Conflictos semióticos en el Cálculo del p -valor

Conflicto	Categorías	Frecuencia	Porcentaje
<i>Conceptual</i>			
1. No definir la distribución muestral como una distribución condicionada por el valor hipotético	CC3	9	17,7
2. Definición incorrecta del p -valor como probabilidad puntual	CC2, CC4	6	11,7
<i>Procedimental</i>			
1. Uso incorrecto de la tabla normal	CC2, CC3, CC4	15	29,4
2. Operar mezclando probabilidades simples y condicionales	CC2, CC4	6	11,7
<i>Representacional</i>			
1. No asociar los objetos matemáticos y su terminología	CC2, CC3, CC4	15	29,4

4.5.5. Interpretación de los resultados

Una vez finalizado los cálculos, se espera que el futuro profesor concluya que el resultado es estadísticamente significativo y, como consecuencia, la hipótesis nula debe ser rechazada. Contextualizando este resultado, la conclusión que se debe adoptar es rechazar la hipótesis de que la esperanza media de vida en Europa sea menor o igual que en el conjunto de países.

En el problema planteado, los futuros profesores han de recorrer un ciclo completo de modelización, descrito por Chaput, Girard y Henry (2011): en primer lugar, han de abstraer la realidad (esperanza de vida y los datos del problema) para plantear,

posteriormente, las hipótesis y de ahí pasar a un modelo matemático adecuado (con cualquiera de los procedimientos descritos: test de Fisher, contraste de Neyman-Pearson, intervalo de confianza o máxima verosimilitud). El tercer paso es trabajar con el modelo matemático elegido para obtener una conclusión y finalizar traduciendo dicha conclusión a la realidad estudiada (esperanza de vida media al nacer).

El proceso de interpretación de los resultados es también parte del proceso de razonamiento estadístico que Wild y Pfannkuch (1999) definen como integración de la estadística y el contexto. En este tipo de proceso el analista ha de navegar entre la esfera del conocimiento estadístico y del conocimiento contextual utilizando ambas para obtener sus conclusiones. Esta importancia nos ha llevado a analizar la capacidad de interpretación y contextualización de los futuros profesores de la muestra, capacidad algo deficiente en futuros profesores (Arteaga, 2011). Considerando tanto la decisión tomada como la contextualización de la misma se han definido nueve las siguientes categorías.

CI1. *Decisión correcta y contextualización adecuada.* Engloba aquellos casos en los que el futuro profesor toma la decisión correcta, rechazando la hipótesis nula y, además, contextualiza el significado que este rechazo tiene en el contexto del problema. Para ello argumenta que se debe rechazar que la esperanza media de vida en Europa sea igual o menor a la del resto de países en el estudio, como queda reflejado en la respuesta siguiente:

Comparando este valor tan pequeño con el nivel de significación $\alpha=0.05$, lo más razonable sería rechazar la hipótesis nula y concluir que la esperanza de vida media en Europa es mayor que en el conjunto de países (Respuesta de GPF).

CI2. *Decisión y contextualización correcta, pero interpretación del resultado en términos de probabilidad de la hipótesis.* Algunos futuros profesores toman la decisión correcta e incluso son capaces de contextualizarla, mostrando que han completado el ciclo de modelización descrito por Chaput et al. (2011) e integran estadística y contexto. Sin embargo, realizan una interpretación errónea del p -valor, ya que lo consideran como la probabilidad de que la hipótesis finalmente aceptada sea correcta, es decir, intercambian los términos en la definición del p -valor. Estos futuros profesores estarían dando una interpretación bayesiana a los resultados del contraste, lo cual es inadecuado para el procedimiento de cálculo empleado. Esta lógica híbrida fue descrita,

entre otros, por Birnbaum (1982), Falk (1986), Gigerenzer (1993) y Vallecillos (1994). Este conflicto se observa en la respuesta dada por BCM.

En resumen, la esperanza media de vida en Europa es mayor que la esperanza media de vida en el conjunto de países, con una probabilidad de 0,9998 (Respuesta de BCM).

CI3. *Decisión y contextualización correctas, pero se indica que la hipótesis nula es falsa.* En este caso, aunque la decisión y contextualización es correcta, se muestra una concepción determinista del resultado del contraste de hipótesis, asumiendo que la hipótesis nula es falsa, lo cual supone un conflicto semiótico conceptual. Dicha concepción errónea, descrita por Inzunsa y Jiménez (2013) y Vallecillos (1994), implica falta de comprensión de la lógica del contraste de hipótesis, como se muestra en la siguiente respuesta:

El valor crítico de z_{α} con nivel de significación 0,05 es igual a 1,645, y dado que $\frac{78-69,2}{2,5} = 3,52$ es superior concluimos que la hipótesis nula ha de ser falsa, así que aceptamos la alternativa que es que la media de la esperanza de vida en Europa es superior a la del total de países a un nivel de significación 0,05 (Respuesta de ALP).

CI4. *Decisión correcta con contextualización inapropiada.* Cuando se indica que la hipótesis nula debe ser rechazada, decisión no en consonancia con la conclusión contextualizada. En consecuencia, no se es capaz de integrar la estadística con el contexto (Wild y Pfannkuch, 1999) o de finalizar el ciclo de modelización (Chaput et al., 2011). Como ejemplo de ella se muestra la siguiente respuesta:

Por tanto, obtenemos un p-valor igual a 0,001, que es menor que el nivel de significación fijado. Concluimos rechazando la hipótesis nula, por lo que la media de los países de la muestra no puede ser mayor que la media poblacional (Respuesta de ILG).

CI5. *Decisión correcta, pero no contextualizada.* Se incluyen en esta categoría las respuestas en las que se concluye indicando el rechazo de la hipótesis, sin asociar dicha respuesta con el contexto en el que se desarrolla el ejercicio. Tampoco en este caso se es capaz de integrar la estadística con el contexto (Wild y Pfannkuch, 1999) o de finalizar el ciclo de modelización (Chaput et al., 2011):

Pasamos a ver las tablas, en este caso la tabla de distribución normal estándar y para $Z=3,52$ nos da 0,99998 que restándolo a 1 nos da una probabilidad de 0,0002 y por tanto rechazamos la hipótesis (Respuesta de AGS).

CI6. *Indica cómo tomaría la decisión y contextualiza.* Se codifican en esta categoría las respuestas de futuros profesores que no han finalizado el procedimiento y muestran un conflicto procedimental. Así, en el ejemplo siguiente se observa como el participante, que optó por el procedimiento de intervalo de confianza, no llega a calcular los valores numéricos del extremo, pero sí que realiza una contextualización adecuada.

Nuestra variable será la esperanza de vida y nuestra hipótesis nula que la media es de 78, como no nos sale dentro del intervalo, la rechazamos, por lo que estimamos que la esperanza de vida en Europa es mayor que en el conjunto de países que hemos escogido (Respuesta de EGM).

CI7. *Indica cómo tomaría la decisión, pero sin contextualizar.* Este caso suele corresponder a futuros profesores que no han finalizado el cálculo del valor crítico o del intervalo de confianza correspondiente. En el siguiente ejemplo, se observa como el participante describe correctamente el método, pero sin llegar a calcular los valores numéricos del extremo.

Este estadístico, llamado estadístico experimental, nos da la probabilidad de encontrar un valor en la distribución normal $N(0,1)$ con una desviación típica de 3,52. Comparando el valor que encontremos en una tabla de probabilidades de una distribución normal con el valor de nuestro contraste hipótesis, podremos determinar si rechazamos la hipótesis nula o la retenemos (Respuesta de MMR).

CI8. *Decisión incorrecta.* A continuación, incluimos una respuesta que basa su procedimiento en el intervalo de confianza y mediante el cual se concluye que la hipótesis nula debe ser rechazada en el caso de que el valor supuesto del parámetro no está incluido en el intervalo construido. Además de concluir incorrectamente, se observa que el resultado tomado no es contextualizado.

H_0 = la esperanza de vida media en Europa es mayor que la esperanza de vida media en el resto de países. [76,78; 79,23]. $H_0 \rightarrow$ No se rechaza, ya que el valor medio del conjunto de países queda por debajo de los valores del intervalo de confianza para un nivel de confianza del 95% de la muestra de países europeos (Respuesta de JGM).

CI9. *No toma decisión.* No se indica cómo y cuál es la decisión del contraste, como el caso que sigue, donde, entre varios errores, no se refleja la decisión a tomar:

Tipificamos $P = \frac{78-69,2}{2,5} \mu_2 = 69,2 P(Z \geq 3,5)$ (Respuesta de AGP).

En la Tabla 4.5.9 se presentan los resultados de la interpretación realizada por los futuros profesores, donde menos del 40% toman la decisión correcta,

contextualizando el problema planteado; aunque otro 47,9% (categorías CI2, CI3, CI4 y CI5) toma la decisión correcta de rechazar la hipótesis nula, pero con algún error. A pesar de la importancia que tiene la contextualización, pues denota una comprensión profunda del problema y del interés del contraste de hipótesis, se ha observado que el 41,1% presentan dificultades a la hora de asociar el resultado de rechazar la hipótesis nula con significado en el problema que se está tratando, es decir, de desarrollar el último paso del proceso de modelización descrito por Chaput et al. (2011). Este último paso de contextualización fue también difícil en la investigación de Arteaga (2011), donde se pidió a futuros profesores de educación primaria trabajar en un proyecto estadístico. El resto de futuros profesores o bien no toman una decisión (aunque describan cómo tomarla) o la toman en forma incorrecta. Además, algunos muestran errores de comprensión del contraste de hipótesis, asumiendo que el p -valor es la probabilidad de que la hipótesis nula sea correcta o adoptando una conclusión determinista de los resultados.

Tabla 4.5.9.

Distribución de Respuestas en la Interpretación de los Resultados

Código	Interpretación de los resultados	Frecuencia	Porcentaje
CI1	Decisión correcta y contextualización adecuada	28	38,3
CI2	Decisión y contextualización correcta, pero interpretación del resultado en términos de probabilidad de la hipótesis	3	4,1
CI3	Decisión y contextualización correctas, pero se indica que la hipótesis nula es falsa	2	2,7
CI4	Decisión correcta con contextualización inapropiada	6	8,2
CI5	Decisión correcta, pero no contextualizada	24	32,9
CI6	Indica cómo tomaría la decisión y contextualiza	1	1,4
CI7	Indica cómo tomaría la decisión, pero sin contextualizar	7	9,6
CI8	Decisión incorrecta	1	1,4
CI9	No toma decisión	1	1,4
Total		73	100,0

En la Tabla 4.5.10 se presentan los conflictos semióticos asociados a la interpretación de los resultados. Vemos que la frecuencia es pequeña, por tanto, aunque pocos estudiantes llegan a interpretar y contextualizar, no se observan muchos conflictos semióticos en este apartado.

Tabla 4.5.10.

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en la Interpretación de los Resultados

Conflicto	Categorías	Frecuencia	Porcentaje
<i>Conceptual</i>			
1. Intercambiar los términos de la probabilidad condicional en la definición del p -valor	CI2	3	4,1
2. Concepción determinista del contraste de hipótesis	CI3	2	2,7
<i>Procedimental</i>			
1. Cálculo erróneo de la región crítica	CI6, CI7	8	11,0
2. Decisión incorrecta o falta de decisión en el contraste de hipótesis	CI8, CI9	2	2,8

4.5.6. Síntesis de conocimientos sobre el contraste de hipótesis

Con objeto de analizar el conocimiento que tienen los futuros profesores sobre el contraste de hipótesis se han contabilizado el número de respuestas correctas dadas por los mismos a cada uno de los puntos descritos a lo largo de la presente sección (el número máximo de apartados correctos será de cinco). Un conocimiento adecuado sobre el contraste de hipótesis supone que el futuro profesor ha descrito al menos cuatro categorías correctas (nótese que la determinación del p -valor únicamente se ha analizado en las respuestas de 51 participantes). Se consideran respuestas parcialmente correctas aquellas que tienen tres apartados correctos e incorrectos los restantes.

En la Tabla 4.5.11 se presentan los resultados sobre el número total de apartados correctos por cada participante. Se observa que menos de un cuarto de los participantes muestran un correcto conocimiento sobre el tema en cuestión y solamente uno de ellos tiene una calificación de cinco puntos. Obsérvese que casi la mitad de los participantes han resuelto el problema incorrectamente ya que no han descrito al menos tres apartados correctamente; señalamos que seis no han desarrollado adecuadamente ninguna de las etapas en la que se ha dividido el desarrollo del contraste.

La mayor parte de los futuros profesores (32%) han desarrollado correctamente tres de los apartados analizados. En este caso particular, el análisis realizado nos ha permitido observar que aproximadamente la mitad de los participantes han planteado las hipótesis del contraste desarrollando un procedimiento adecuado y han obtenido un p -valor de 0,00135 (11 participantes) o han realizado una interpretación correcta de los resultados obtenidos (7 participantes). Asimismo, se destaca que tres de los participantes han llevado a cabo un procedimiento adecuado, un cálculo correcto del p -valor y han realizado una interpretación correcta los resultados obtenidos tras la realización del contraste de hipótesis, resultado llamativo pues el futuro profesor no ha sido capaz de determinar correctamente las hipótesis del contraste.

Tabla 4.5.11.

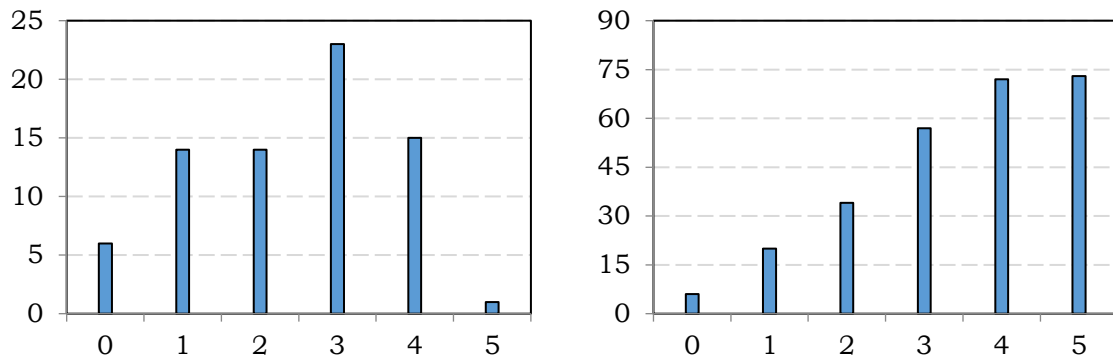
Distribución del Número de Apartados Correctos Sobre el Contraste de Hipótesis

Apartados correctos	Frecuencia	Porcentaje	Fr. Acumulada	% Acumulado
0	6	8,2	6	8,2
1	14	19,2	20	27,4
2	14	19,2	34	46,6
3	23	31,5	57	78,1
4	15	20,5	72	98,6
5	1	1,4	73	100

La representación gráfica de la información recogida en la Tabla 4.5.11 evidencia que la mayoría de los participantes han tenido tres apartados correctos (véase Figura 4.5.1). Asimismo, se observa en la gráfica de la frecuencia acumulada un crecimiento considerable hasta el valor modal (3) y para los dos últimos valores este crecimiento es casi despreciable por el aumento tan pequeño que supone.

De los resultados obtenidos se pone de manifiesto las dificultades que presenta para los futuros profesores de Educación Secundaria Obligatoria y Bachillerato, participantes en este estudio, que podrían ser compartida con otros futuros profesores, al abordar un problema de contraste de hipótesis. Esta conclusión se refuerza por la variedad de conflictos semióticos que se han identificado en cada uno de los pasos del proceso y que se compilan en la en la Tabla 4.5.12.

Figura 4.5.1. Representación gráfica de la distribución y distribución acumulada del número de apartados correctos por participante en el contraste de hipótesis



Estas conclusiones tienen que ser tenidas en cuenta en la preparación de los futuros profesores ya que ellos son los encargados de formar a los estudiantes en dicho tema. Es evidente que si los futuros profesores no tienen un conocimiento adecuado sobre el contraste de hipótesis difícilmente transmitirán correctamente a sus estudiantes cómo deben afrontar la elaboración de problemas similares al presentado en este estudio.

4.5.7. Conclusiones sobre el conocimiento común del contraste de hipótesis

Para finalizar el estudio del primer problema planteado, resumimos las principales conclusiones obtenida del análisis de las soluciones al mismo. Lo más sencillo en este problema fue el planteamiento de las hipótesis, ya que la mayoría de los participantes las formularon correctamente.

También en su mayoría eligen un procedimiento consistente con las hipótesis planteadas, aunque el 84% no declaran explícitamente el tipo de contraste elegido y otros confunden el tipo de contraste, por ejemplo, confunde contraste bilateral y unilateral, como ocurre en el trabajo de Vallecillos (1994). Son pocos los que desarrollan correctamente todos los pasos del contraste elegido (15,1%) y algunos cometen también errores o imprecisiones en el cálculo del p -valor.

Tabla 4.5.12.

Conflictos semióticos de los futuros profesores sobre el contraste de hipótesis

Conceptuales

1. Considerar hipótesis no complementarias en el contraste de hipótesis (CH4, CH11)
2. No comprender el papel del valor puntual de la hipótesis nula en la elección de la distribución muestral del estadístico (CH4)
3. Confundir la hipótesis nula y alternativa, y sus papeles en el contraste (CH5, CH8, CP2)
4. Confundir estadístico y parámetro (CH5, CH6, CH8, CP2)
5. No comprender el papel del valor supuesto puntual en la hipótesis nula, asignando el valor puntual a la hipótesis alternativa (CH10, CH11)
6. Confunde un contraste de hipótesis unilateral derecho con otro unilateral izquierdo (CP2)
7. Confundir un contraste de hipótesis unilateral con otro bilateral (CP5, CD2)
8. Confundir región crítica unilateral y bilateral (CP5)
9. Confundir región crítica y de aceptación (CD7)
10. No definir la distribución muestral como una distribución condicionada por el valor hipotético (CP4, CD4, CC3)
11. Confundir las condiciones en que se puede aplicar una distribución normal para la distribución de la media muestral (CD3)
12. Definir el p -valor, sin incluir en su cálculo el valor hipotético (CD5)
13. Definición incorrecta del p -valor como probabilidad puntual (CD6, CC2, CC4)
14. Intercambiar los términos de la probabilidad condicional en la definición del p -valor (CI2).
15. Confundir variable aleatoria continua y discreta (CH9)
16. No comprender que la probabilidad puntual de una variable aleatoria continua es siempre igual a cero (CD6)
17. Concepción determinista del contraste de hipótesis (CI3)

Procedimentales

1. Falta de comprensión o incorrecta aplicación de la lógica del contraste de hipótesis (CH5, CH6)
2. Olvida plantear una de las dos hipótesis en un contraste (CH7, CH8)
3. Cálculo erróneo de la región crítica (CD2, CI6, CI7)
4. Cálculo erróneo de intervalo de confianza (CD2)
5. Cálculo incorrecto de valor crítico (CD3)
6. Comparar valores críticos procedentes de diferentes distribuciones (CD3)
7. Aplicar erróneamente el resultado de la razón de verosimilitudes (CD7)
8. Uso incorrecto de la tabla normal (CC2, CC3, CC4)
9. Operar con mezcla de probabilidades simples y condicionales (CC2, CC4)
10. Cálculo erróneo de la región crítica (CI6, CI7)
11. Decisión incorrecta o falta de decisión en el contraste de hipótesis (CI8, CI9)

Representacionales:

1. No asociar los objetos matemáticos y los símbolos que los representan (CH2, CH3, CD4, CC2, CC3, CC4)
 2. No recordar la terminología que corresponde a algunos objetos matemáticos (CP3, CP4)
-

Los errores detectados en este trabajo reproducen los identificados en estudiantes en investigaciones previas, como las de Cañadas et al. (2012), Vallecillos (1994), y plantean la pregunta de hasta qué punto dichos errores podrían ser

transmitidos en la enseñanza, si el futuro profesor no cuenta con un conocimiento suficiente del tema. En este trabajo los hemos interpretado en términos de conflictos semióticos, clasificándolos en conceptuales, procedimentales o representacionales.

Además, se han encontrado algunos no descritos anteriormente, como no comprender el papel del valor supuesto puntual en la hipótesis nula, asignando el valor puntual a la hipótesis alternativa.

Por otro lado, aunque la mayoría de los participantes tomó una decisión correcta en el contraste planteado (86.2%) , fueron pocos los que tomaron una decisión correcta y llegaron a interpretar este resultado en el contexto del problema (38,7%). Al igual que ocurrió en el trabajo de Arteaga (2011) con futuros profesores de primaria, parece que los participantes ven como único objetivo del problema el desarrollo de los cálculos y no los interpretan. Por tanto, llegan a sugerir rechazar la hipótesis, sin indicar cuál es esta hipótesis o que lo que se rechaza es la idea de que la esperanza de vida al nacer en Europa sea la misma que la calculada, teniendo en cuenta todos los países de la muestra. Los resultados obtenidos en relación a esta variable son inquietantes ya que se espera que los futuros profesores desarrollen en los estudiantes la capacidad de realizar interpretaciones de estudios estadísticos tomando como referencia contextos cercanos al estudiante (MECD, 2015a, p. 398 y p. 407).

Los resultados obtenidos nos advierten de la necesidad de preparar al futuro profesorado de Educación Secundaria Obligatoria y Bachillerato en el tema de inferencia estadística, y más concretamente en el contraste de hipótesis. No sólo este contenido ha sido tratado en las directrices curriculares y pruebas de acceso a la universidad del Bachillerato de Humanidades y Ciencias Sociales hasta el 2015, sino que es un tema de las oposiciones de estos profesores. Por otro lado, la lectura crítica e interpretación de las investigaciones en educación matemática, que debieran analizar estos profesionales para mantenerse actualizados sobre los avances de didáctica de la matemática, requiere un conocimiento suficiente del tema, pues estas investigaciones con frecuencia incluyen resultados de contrastes de hipótesis.

Es por tanto necesario, una mayor formación estadística de los futuros profesores, incluyendo experiencias como la descrita por Dolor y Noll (2015) con aproximaciones informales a la inferencia y contraste de hipótesis sobre variables categóricas. Es importante también reforzar su conocimiento de las distribuciones

muestrales (Batanero, 2000; Liu y Thompson, 2005), pues algunos de los conflictos semióticos detectados se relacionan con su comprensión.

4.6. CONOCIMIENTO MATEMÁTICO SOBRE EL INTERVALO DE CONFIANZA

Para evaluar el conocimiento matemático común de los futuros profesores sobre el intervalo de confianza se les propuso que resolvieran detalladamente el ítem 2 presentado en la Subsección 4.4.2. Recordemos que el objeto del mismo es la determinación de un intervalo de confianza al 95% para la esperanza media de vida en los países europeos empelando para ello la información recogida en el ítem 1. La elaboración del ítem supone además que los participantes deberán indicar cómo explicaría a un estudiante tanto el significado como algunas de las propiedades fundamentales del intervalo de confianza.

Respecto al análisis de este ítem, se ha descompuesto la solución aportada por el estudiante en varios apartados. En primer lugar, se tienen en cuenta tres puntos para la primera parte del ítem, donde se pide construir el intervalo:

1. Planteamiento inicial para la construcción del intervalo, es decir, forma en que se propone dicha construcción.
2. Cálculo del intervalo. Operaciones realizadas para obtener los extremos y su corrección.
3. Interpretación del intervalo de confianza obtenido.

Puesto que, además, en la segunda parte se pregunta la forma en que se explicaría a un estudiante el intervalo de confianza y sus propiedades, también se consideran los siguientes puntos:

4. Definición de intervalo de confianza. Forma en que definirían el intervalo a un estudiante.
5. Propiedades que se recuerdan o se explicarían a un estudiante sobre los intervalos de confianza.

A continuación, desarrollamos cada uno de estos puntos acompañándolos con ejemplos de respuestas dadas por los futuros profesores.

4.6.1. Planteamiento del intervalo de confianza

En este apartado analizamos la forma en que el estudiante plantea la construcción del intervalo de confianza. Para llevarlo a cabo, en primer lugar, debe recordarse que si se consideran todas las posibles muestras de tamaño n , que pueden extraerse de una población determinada, cada una de ellas tendrá un valor de media distinto. Dado que dicho valor varía de una muestra a otra, entonces, la media muestral, denotada por \bar{X} , adquiere el carácter de variable aleatoria (Schuyten, 1991) y, como consecuencia, se distribuirá según un modelo probabilístico determinado.

En base al enunciado del ítem 1, donde se supone que la distribución de partida es normal, y en base a la propiedad reproductiva respecto a sus parámetros, se verifica que la media muestral sigue una distribución normal, pues se define como suma de variables normales, salvo una constante ($1/n$). Aplicando las propiedades de linealidad de la media y de la varianza (en este caso se puede aplicar porque las variables sumadas son independientes), se obtiene para la media de la muestra una distribución normal de media μ (igual que la media de la población de donde se ha extraído) y desviación típica σ/\sqrt{n} (cociente entre la desviación típica de la población y la raíz cuadrada del tamaño de la muestra extraída), es decir, $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$. La tipificación de la variable, $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, implica la ventaja de poder hacer uso de las tablas estadísticas de una distribución normal estandarizada, $N(0,1)$.

La construcción del intervalo de confianza se centra en encontrar los límites de un intervalo de valores (e_1, e_2) , con un nivel de confianza $(1 - \alpha) \times 100\%$, de tal forma que se verifique que $P(e_1 < p < e_2) = 1 - \alpha$, siendo p el parámetro poblacional que se desea estimar (en nuestro caso el parámetro a estimar es la media poblacional, denotada por μ). El valor $1 - \alpha$ o nivel de confianza se interpreta como probabilidad del proceso, es decir, en 100 intervalos construidos a partir de diferentes muestras de tamaño n extraídas de la misma población, el $(1 - \alpha) \times 100\%$ aproximadamente cubrirá el valor del parámetro. En relación a la variable *planteamiento*, se han encontrado las siguientes cuatro respuestas:

CIP1. *Respuesta correcta.* Se especifica que el objetivo es hallar los valores de a y b que cumplen que $P(a < \mu < b) = 0,95$ o que explicitan de algún otro modo correcto la

expresión del intervalo de confianza. Entre las respuestas más completas destacamos la dada por el participante BDM, donde, además de describir las características de las variables en estudio, realiza de forma precisa y correcta el desarrollo teórico que da como resultado la expresión del intervalo de confianza para la media de una población normal con varianza conocida. Observamos, no obstante, que el estudiante desarrolla todo el procedimiento, pero no explica en base a qué decide que la distribución muestral es normal o cuáles son sus parámetros.

Para hacer el intervalo de confianza de una variable que sigue una distribución normal cuya media es 69,2 y desviación típica 10, esto es, $X \sim N(69,2; 10)$, se utiliza el pivote $T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$. Entonces $P(\lambda_1 < T < \lambda_2) = P\left(\lambda_1 < \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \lambda_2\right) = P\left(\bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}}\right) = P(\lambda'_1 < \mu_0 < \lambda'_2)$. Siendo $P(\lambda'_1 < \mu_0 < \lambda'_2) = \phi(\lambda'_2) - \phi(\lambda'_1)$ que, como tiene que ser un intervalo de confianza del 95%, entonces $\phi(\lambda'_2) - \phi(\lambda'_1) = 1 - \alpha = 0,95 \Rightarrow (\lambda'_1, \lambda'_2) = \left(\bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}}, \bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}}\right)$. Por tanto, $\lambda_2 = -\lambda_1$ y $\lambda_1 = z_{\frac{\alpha}{2}}$ con $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$. Entonces el intervalo de confianza del 95% será $\left(\bar{X} - \lambda_2 \frac{\sigma}{\sqrt{n}}, \bar{X} - \lambda_1 \frac{\sigma}{\sqrt{n}}\right)$ (Respuesta de BDM).

CIP2. *Explicar lo que es el intervalo sin planteamiento.* Se contemplan aquellas respuestas donde el participante describe verbalmente la forma en que se plantea la construcción de un intervalo de confianza, sin llegar a definir la expresión asociada a éste, pues posiblemente no la recuerda. Obsérvese que la respuesta dada por SS no deja entrever si el participante sabe plantear correctamente lo que se solicita en el problema por lo que subyace un conflicto semiótico procedimental, consistente en no recordar la expresión de cálculo del intervalo de confianza. Además, la expresión “varios pares de números” no deja claro si se refiere a un único intervalo, pues este sería únicamente un par de valores.

Otro error en la respuesta es suponer que el complementario del coeficiente de confianza α es una probabilidad de error en la estimación, interpretación incorrecta similar a otras dadas sobre el nivel de significación de un contraste, por ejemplo, en Birnbaum (1982), Falk (1986) y Vallecillos (1994) y que supone un conflicto semiótico. En realidad, el nivel de confianza nos da la probabilidad de que el método construya un intervalo que incluya al parámetro desconocido, es decir, se refiere al procedimiento y no al resultado del mismo (Garfield, delMas y Chance, 2004),

El intervalo de confianza es un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto. Estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. La probabilidad de éxito en la estimación

se representa con $1 - \alpha$ y se denomina “nivel de confianza”. En estas circunstancias, α es el llamado “error aleatorio” o “nivel de significación”, esto es, una medida de las posibilidades de fallar en la estimación mediante tal intervalo (Respuesta de SS).

CIP3. *Plantear el intervalo usando la media poblacional en lugar de la media muestral.* En esta categoría de respuesta, se describe la expresión del intervalo utilizando la media poblacional en lugar de la media muestral o donde emplea el valor 69,2 (valor de la media poblacional) con el fin de obtener los extremos del intervalo. Esta respuesta es indicativa de la confusión entre estadístico y parámetro es un conflicto semiótico que también apareció en el contraste de hipótesis y ha sido detectada en investigaciones previas sobre inferencia estadística (Harradine et al., 2011; Schuyten, 2001; Vallecillos, 1999). En el caso del intervalo de confianza, Garfield et al. (2004) sugieren que muchos estudiantes piensan que el intervalo contiene al estadístico, en vez de al parámetro.

A modo de ejemplo se muestra la respuesta del participante MTF donde se construye un intervalo centrado en la media supuesta de la población, dada en el ítem 1, y no en la media obtenida en la muestra de países europeos. Por otro lado, el estudiante da una interpretación determinista al intervalo, pues no aparece en su respuesta ninguna expresión de probabilidad o de confianza, lo que sería otro conflicto semiótico. El estudiante no comprende la finalidad de la estimación de los parámetros, cuando indica que se puede comprobar si el parámetro pertenece al intervalo.

Lo que tenemos que hacer es usar la fórmula que hay para intervalos de confianza para la media. Y una vez que tenemos la fórmula sustituir y comprobar si el valor que queremos comprobar pertenece al intervalo o no, si pertenece podemos decir que la esperanza de vida en los países europeos es la que se quiere comprobar. $(69,2 - 1,96 \times \frac{10}{\sqrt{16}}; 69,2 + 1,96 \times \frac{10}{\sqrt{16}})$ (Respuesta de MTF).

CIP4. *Planteamiento incorrecto de la fórmula del intervalo.* En esta categoría se han considerado aquellos casos en los que el participante, aunque emplea la notación adecuada, no describe correctamente la expresión del intervalo, pues olvida o confunde alguno de los términos que intervienen en el cálculo. Obsérvese por ejemplo la respuesta dada por PFS, donde el participante identifica correctamente la distribución de la variable media muestral (denotada por X_m) y señala que el intervalo de confianza es un intervalo donde se estima que estará la media poblacional. Sin embargo, a la hora de dar la expresión correspondiente al intervalo de confianza pedido introduce, en ambos

extremos del intervalo, la media de la muestra precedida de un signo negativo en lugar de un signo positivo. No hemos encontrado este error descrito en la literatura.

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

La Tabla 4.6.1 recoge los resultados obtenidos en referencia a este apartado. Se observa que la mayoría de los participantes en la investigación plantean de una forma correcta la construcción del intervalo, resultados que son bastante buenos. Se han obtenido cinco repuestas imprecisas ya que los participantes describen verbalmente el proceso, pero no describen la expresión de cálculo del mismo, no pudiendo saber si éstos conocen la notación adecuada o la fórmula para obtener el intervalo solicitado. Sólo hay dos expresiones del intervalo claramente incorrectas. Puesto que la mayoría de las resoluciones son correctas, se presentan pocos conflictos semióticos, pero, para continuar el mismo sistema seguido en los apartados anteriores, estos conflictos son descritos en la Tabla 4.6.2.

Tabla 4.6.1.

Distribución de Respuestas en el Planteamiento del Intervalo

Código	Planteamiento	Frecuencia	Porcentaje
CIP1	Respuesta correcta	64	87,7
CIP2	Explicar lo que es el intervalo sin planteamiento	5	6,9
CIP3	Plantear el intervalo usando la media poblacional en lugar de la media muestral	2	2,7
CIP4	Planteamiento incorrecto de la fórmula del intervalo	2	2,7
Total		73	100,0

Respecto a los conflictos semióticos detectados, se ha hallado el citado por Harradine et al. (2011) y Vallecillos (1999) en relación a la confusión entre estadísticos muestrales y parámetros poblacionales y en el uso incorrecto de la expresión asociada al intervalo. Igualmente aparecen concepciones deterministas y definiciones incorrectas del intervalo. Estos conflictos tienen muy baja frecuencia, debido a que el planteamiento del intervalo fue generalmente correcto.

4.6.2. Cálculo del intervalo

Una vez definido o planteado el intervalo de confianza, se deben identificar los datos del enunciado del problema para posteriormente obtener los valores numéricos que determinan los extremos del este. Como ya se ha indicado, el cálculo del intervalo de confianza supone la identificación de los valores de una distribución muestral normal y su distribución normal tipificada centrada en el origen, que depende de su media y desviación típica. Además, se deben determinar los puntos críticos, a partir de la lectura de la tabla de la distribución normal $N(0, 1)$, de tal manera que el área de cada cola sea igual a $\alpha/2$, es decir, $P(-k < Z < k) = 1 - \alpha$ donde $Z \sim N(0, 1)$.

Tabla 4.6.2.

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en el Planteamiento del Intervalo

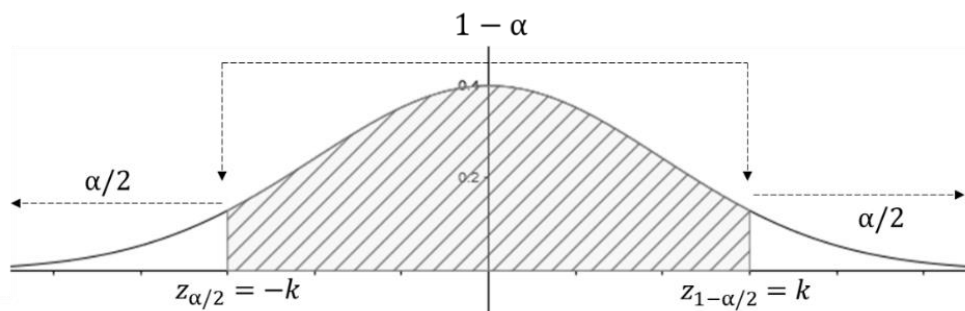
Conflicto	Categorías	Frecuencia	Porcentaje
<i>Conceptual</i>			
1. Confundir estadístico (media muestral) y parámetro (media poblacional)	CIP3	2	2,7
2. Definir el coeficiente de confianza o su complementario como probabilidad de error	CIP2	5	6,9
3. Concepción determinista del intervalo de confianza	CIP3	2	2,7
4. Definir el intervalo como “varios pares de números”	CIP2	5	6,9
5. Pensar que se puede comprobar que el intervalo contiene al parámetro	CIP3	2	2,7
<i>Procedimental</i>			
1. No recordar la expresión de cálculo del intervalo de confianza	CIP2	5	6,9
2. Error en la fórmula de cálculo de los extremos del intervalo	CIP4	2	2,7

Realizando una lectura inversa de la tabla de la distribución estandarizada, se tiene que los valores $-k$ y k que cumplen dicha condición son aquellos cuyas probabilidades son $P(Z < -k) = \alpha/2$ y $P(Z < k) = 1 - \alpha/2$, respectivamente (véase Figura 4.6.1). En este caso, dado que el nivel de confianza considerado es del 95%, entonces se tiene que $-k = Z_{\alpha/2} = Z_{0,025} = -1,96$ y $k = Z_{1-\alpha/2} = Z_{0,975} = 1,96$. Puesto que la variable asociada al estadístico muestral tipificado $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ se distribuye según $N(0; 1)$, se tiene, sustituyendo en $P(-k < Z < k) = 1 - \alpha$ y despejando el parámetro poblacional, el intervalo de confianza para la media poblacional y cuya expresión viene dada por

$$\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right).$$

Sustituyendo la información muestral ($\bar{X} = 78, n = 16$) y el valor de la desviación típica poblacional ($\sigma = 5$) se obtiene que el intervalo de confianza solicitado es (73,1; 82,9). Teniendo en cuenta el procedimiento llevado a cabo se han definido seis categorías para clasificar el cálculo realizado por los futuros profesores y que se detallan a continuación.

Figura 4.6.1. Selección de los Puntos Críticos (para una Distribución Normal Tipificada) en el Cálculo del Intervalo de Confianza



CIC1. *Cálculo correcto del intervalo de confianza.* Se consideran aquellas respuestas en las que los participantes obtienen los valores de los extremos del intervalo pedido, realizando todos los pasos que se han descrito y sin cometer errores. A modo de ejemplo, se muestra la respuesta dada por AMC donde se observa la correcta obtención de los dos cuantiles en la distribución normal tipificada, y la elección adecuada de los valores de cada uno de los elementos que determinan la expresión del intervalo, es decir, media muestral y varianza de la distribución muestral, así como valores críticos y cálculo correcto de los extremos, realizando las operaciones requeridas.

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

CIC2. *Aproximar el valor de la distribución normal, obteniendo un intervalo aproximado.* Algunos futuros profesores realizan la mayor parte de los pasos descritos en la categoría CIC1 correctamente, pero en lugar de utilizar los valores exactos de los cuantiles de la distribución normal $N(0,1)$ o valores críticos, los aproximan, utilizando la propiedad de los intervalos de la distribución normal que contienen un cierto porcentaje de casos.

Al igual que ocurrió en el contraste de hipótesis, posiblemente se deba al olvido de la forma en que se leen las tablas de la distribución normal. En el ejemplo siguiente, el participante aproxima por 2 el valor de la distribución normal que comprende, aproximadamente, el 95% de los datos y por ello obtiene un valor aproximado para los extremos del intervalo. Por otro lado, este futuro profesor no llega a sustituir la expresión por el valor de Z para determinar los extremos, por lo que muestra dificultad para el trabajo con desigualdades.

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

CIC3. *Pequeños errores de cálculo en las operaciones.* Se clasifica en esta categoría la respuesta del estudiante, cuando todo el planteamiento de la construcción del intervalo de confianza es correcto, pero se cometen errores en el cálculo de uno de los extremos. Un ejemplo lo tenemos en la respuesta de LAP que comete un error en el cálculo del extremo superior del intervalo, obteniendo 85,84 en lugar de 82,9:

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

CIC4. *Explicar cómo se calcularía el intervalo de confianza sin realizar los cálculos.* Se consideran aquellos casos en los que se incluye un planteamiento adecuado de la construcción del intervalo, pero sin llegar a obtener los valores que determinan los extremos del intervalo. Ejemplo de ello es la respuesta dada por JBL donde se evidencia que el participante tiene conocimientos sobre la construcción del intervalo, pero no llega a concluir lo solicitado en el ítem. Posiblemente ha tenido dificultad en identificar los datos del enunciado o lectura de las tablas de la distribución normal

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

CIC5. *Utilizar la media poblacional.* Este tipo de error es consecuencia directa de utilizar la media poblacional en lugar de la media muestral en la expresión del intervalo de confianza. La confusión entre estadístico y parámetro ya apareció en el contraste de hipótesis y ha sido detectada en investigaciones previas sobre inferencia estadística (Harradine et al., 2011; Schuyten, 2011; Vallecillos, 1999). A continuación, se muestra la respuesta dada por AML.

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

En la Tabla 4.6.3 se presenta la frecuencia y porcentaje de futuros profesores que han sido clasificados en cada una de las categorías respecto a la construcción del intervalo de confianza. Se observa que más de la mitad de los participantes dan como resultado el intervalo correcto (73,1; 82,9) y un quinto explica cómo se obtiene el intervalo de confianza, sin llegar a realizar cálculo alguno. En todo caso, son menos los que son capaces de calcular correctamente el intervalo de confianza que aquellos que lo plantean correctamente. Un 38,4% de participantes en el estudio comete algún tipo de error en el cálculo, lo que nos parece muy alto, teniendo en cuenta que se trata de futuros profesores que han de explicar el tema.

Tabla 4.6.3.

Distribución de Cálculos Desarrollados para Obtener el Intervalo de Confianza

Código	Respuesta	Frecuencia	Porcentaje
CIC1	Correcto	45	61,6
CIC2	Aproxima el valor de la distribución normal	5	6,9
CIC3	Pequeños errores de cálculo en las operaciones	4	5,5
CIC4	Explica cómo, pero no realiza los cálculos	15	20,5
CIC5	Utiliza la media poblacional	4	5,5
Total		73	100,0

Se observa que se ha duplicado, respecto a los resultados de la categoría CIP3 de la Tabla 4.6.1, el número de futuros profesores que utilizan la media poblacional en

lugar de la media muestral en el cálculo del intervalo. Este hecho muestra un cierto problema, ya que una confusión de este tipo, en el ámbito de la inferencia estadística, supone falta de comprensión de uno de los principales objetivos para los que se desarrolla el entramado teórico de dicha materia. El resto de errores supone un 17,9% y se han dividido casi en la misma proporción entre los participantes. Por último, señalamos que un 20,5% de la muestra no son capaces de completar el cálculo del intervalo.

La Tabla 4.6.4 recoge los conflictos semióticos detectados al realizar el cálculo del intervalo de confianza, que son pocos y con pequeña frecuencia. Lo más común fue el manejo incorrecto de la tabla de la distribución normal, que también se dio en el contraste de hipótesis. Además, hubo cierta dificultad al identificar los datos en el enunciado del problema.

Tabla 4.6.4.

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en el Cálculo del Intervalo

Conflicto	Categorías	Frecuencia	Porcentaje
<i>Conceptual</i>			
1. Confundir estadístico y parámetro	CIC5	4	5,5
<i>Procedimental</i>			
1. Uso incorrecto de la tabla normal	CIC2, CIC4	19	27,4
2. Errores de cálculo	CIC3	4	5,5
3. Dificultad en trabajo con desigualdades	CIC2	5	6,9
4. Dificultad en identificar los datos	CIC4	15	20,5

4.6.3. Interpretación del intervalo

Una vez obtenidos los extremos del intervalo, se espera que los estudiantes den una interpretación del significado del nivel de confianza, 95%, similar a la siguiente: Si se construyeran 100 intervalos con muestras de tamaño 16, aproximadamente en 95 de ellos estará el verdadero valor de la media poblacional. Es decir, se espera que interpreten los extremos del intervalo como variables (Behar, 2001; Olivo, 2008). En este apartado, y una vez analizadas las respuestas de los futuros profesores, se han definido cinco categorías que son desarrolladas a continuación.

CII1. *Interpretación correcta.* Junto a la interpretación dada anteriormente, se acepta como correcta, aunque sea una versión imprecisa, aquella en la que se indique

que el coeficiente de confianza da el porcentaje de intervalos calculados a partir de muestras de igual tamaño de la población que contiene al parámetro, es decir, que el parámetro está entre los límites calculados con una confianza del 95% (Behar, 2001; Olivo, 2008). Esta sería la respuesta que dan algunos de los futuros profesores. A modo de respuesta correcta se muestra la dada por CMM y como imprecisa la de DCC, donde no queda claro si lo interpreta correctamente.

El intervalo de confianza indica en este caso que de 100 intervalos que construyéramos con distintas muestras, solo en el 95 de ellos la media poblacional pertenecería a ese intervalo (Respuesta de CMM).

(73,1; 82,9), que quiere decir que la media de la esperanza de vida en los países europeos se encuentra entre esos valores con una confianza del 95% (Respuesta de DCC).

Asimismo, y dado que una correcta interpretación del intervalo de confianza supone una adecuada interpretación del nivel de confianza, se han tenido en cuenta aquellos casos en los que se recuerda que la probabilidad dada por el nivel de confianza se refiere a los límites del intervalo (véase respuesta de CBF), considerándolos como variables.

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

CII2. *Interpretación determinista.* Es la respuesta de aquellos participantes que indican que el parámetro se encuentra entre los límites del intervalo, en todos los casos, dando una interpretación determinista del resultado. Pensamos que estos casos visualizan el intervalo de confianza como un estadístico descriptivo, como indican Cumming y Fidler (2005) y no aprecian que los extremos del intervalo son aleatorios (Olivo, 2008). Un ejemplo típico es la respuesta de DAM:

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

CII3. *Interpretación del nivel de confianza en términos probabilísticos.* En esta categoría se hace referencia a aquellas respuestas en las que el nivel de confianza es interpretado en términos de probabilidad, es decir, como la probabilidad de que el parámetro poblacional esté dentro del intervalo obtenido. Sería una interpretación bayesiana y no frecuencial del significado del intervalo. Esta última interpretación

incorrecta fue encontrada en el estudio de Behar (2001), Garfield et al. (2004) y Olivo (2008) e implica que se visualizan los extremos del intervalo como constantes (véase respuesta de LUG).

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

CII4. *Intervalo que contiene a la media de la muestra.* También se observan errores a la hora de interpretar el intervalo como aquel que contiene a la media de la muestra y no a la media de la población, como el siguiente ejemplo. Es de nuevo un caso donde se confunde estadístico y parámetro, error detectado por Behar (2001). Garfield et al. (2004) y Olivo (2008):

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

CII5. *No interpreta.* Se han incluido aquellos casos en los que los futuros profesores no han interpretado los resultados obtenidos.

Los resultados recogidos en la Tabla 4.6.5 refleja las dificultades que tienen los participantes a la hora de dar interpretación a la información obtenida del intervalo de confianza, aproximadamente dos quintas partes no han dado interpretación alguna y un tercio han tenido errores en la misma (categorías CII2, CII3 y CII4). Estos resultados ponen de manifiesto las limitaciones que presentan, en relación a la comprensión del intervalo de confianza, las personas que serán las encargadas de preparar a los estudiantes de secundaria y bachillerato. Señalamos que, de los 21 participantes clasificados en la primera categoría, siete han realizado una interpretación correcta y, salvo dos que definen correctamente el nivel de confianza, el resto han sido algo imprecisos puesto que basan su interpretación en que el parámetro está entre los límites calculados con una confianza del 95%.

Los conflictos semióticos detectados en este apartado se presentan en la Tabla 4.6.6. Como en investigaciones previas, aparece con relativa frecuencia la interpretación bayesiana del intervalo de confianza, sugiriendo que el coeficiente de confianza es la

probabilidad de que el intervalo contenga al parámetro. Un conflicto nuevo es interpretar los extremos del intervalo como constantes en las diferentes muestras, e incluso encontramos futuros profesores que suponen que el intervalo se refiere a la media muestral.

Tabla 4.6.5.

Distribución de Respuestas en la Interpretación de los Resultados Obtenidos

Código	Respuesta	Frecuencia	Porcentaje
CII1	Interpretación correcta	21	28,8
CII2	Interpretación determinista	8	11,0
CII3	Interpretación del nivel de confianza en términos probabilísticos	13	17,8
CII4	Intervalo que contiene a la media de la muestra	1	1,3
CII5	No interpreta	30	41,1
Total		73	100

Tabla 4.6.6

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en la Interpretación de los Resultados

Conflicto	Categorías	Frecuencia	Porcentaje
<i>Conceptual</i>			
1. Definir el coeficiente de confianza como probabilidad de que el parámetro esté comprendido en el mismo	CII3	13	17,8
2. Pensar que el intervalo de confianza contiene a la media muestral	CII4	1	1,3
3. Visualizar los extremos del intervalo como constantes	CII3	13	17,8
4. Concepción determinista del intervalo de confianza	CII2	8	11,0

4.6.4. Definición

En esta variable se han analizado todas las respuestas que incluyen la definición de intervalo de confianza y donde los participantes lo definen cuando son preguntados sobre cómo lo explicarían a los estudiantes. Se espera que los participantes definan un intervalo de confianza como un rango de valores para un parámetro poblacional determinado por dos valores extremos, que permite estimar dicho parámetro a un nivel de confianza determinado. Las tres categorías definidas son las que se muestran a continuación.

CID1. *Definición correcta.* Se han considerado correctas las respuestas en las que se menciona los límites del intervalo, el nivel de confianza y se señala que el intervalo construido es para un parámetro poblacional. Por ejemplo, se indica que un intervalo de confianza es un conjunto de valores plausibles del parámetro que pudiera haberse generado a partir de la muestra observada. Véase la respuesta de BMZ, quien alude al concepto de confianza, parámetro y extremos en la definición.

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

CID2. *Definir el intervalo en términos de probabilidad y no de confianza.* Una definición errónea sobre el intervalo de confianza es aquella en la que el nivel de confianza es interpretado como la probabilidad de que el intervalo contenga el valor poblacional. Es decir, interpretan en forma bayesiana el intervalo de confianza, como se ha informado en otras investigaciones (Behar, 2001, Olivo, 2008). Cumming y Fidler (2005), señalan que esta definición es deficiente ya que, para un intervalo de confianza dado no se puede hablar de probabilidades, puesto que, la probabilidad asociada a un intervalo de confianza, independientemente si incluye o no el verdadero valor, es 1 o 0. Morey et al. (2016) denominan *falacia fundamental de la confianza*, el interpretar dicha confianza como probabilidad de que el parámetro esté contenido en el intervalo. A modo de ejemplo se muestra la respuesta de MSM.

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

CID3. *No define.* Señalamos que no todos los estudiantes han definido lo que es un intervalo de confianza, por lo que, ante esta situación, se ha incluido una categoría para tener en cuenta los futuros profesores que no definen el intervalo.

La Tabla 4.6.7 muestra la frecuencia y porcentaje de respuestas dadas por los futuros profesores en relación a la definición del intervalo de confianza. Se destaca que entre los participantes que han dado una definición sobre el intervalo de confianza, más de la mitad han considerado que el valor del parámetro está entre los extremos del intervalo con una probabilidad determinada, dando una interpretación bayesiana al intervalo. Obsérvese que, aunque hay un porcentaje muy elevado de participantes que

plantean y calculan el intervalo de confianza (véase Tablas 4.6.1 y 4.6.3), aproximadamente dos tercios no definen qué es un intervalo de confianza. Por tanto, podemos sospechar que su aprendizaje es simplemente procedimental y hay una falta de comprensión conceptual del tema.

Dados estos resultados se puede pensar que la resolución de problemas similares al analizado en este estudio no supone mucha dificultad para los estudiantes ya que el proceso queda reducido a la aplicación de una fórmula sin significado alguno. Asimismo, se destaca que el 20% de los estudiantes presentan un conflicto semiótico en la definición del intervalo (véase Tabla 4.6.8).

Tabla 4.6.7.

Frecuencia y Porcentaje de Respuestas sobre Definición

Código	Respuesta	Frecuencia	Porcentaje
CID1	Definición correcta	11	15,1
CID2	Definir el intervalo en términos de probabilidad y no de confianza	15	20,5
CID3	No define	47	64,4
Total		73	100

Tabla 4.6.8

Frecuencia y Porcentaje de Estudiantes con Conflictos Semióticos en la Definición del Intervalo

Conflicto	Categorías	Frecuencia	Porcentaje
Conceptuales			
1. Interpretación bayesiana del intervalo	CID2	15	20,5

4.6.5. Propiedades que se recuerdan

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

CIPR1. *Relación entre nivel de confianza y amplitud del intervalo.* Cuando el futuro profesor cita esta relación, es decir, indica que un aumento (disminución) del nivel de confianza supone un aumento (disminución) en la amplitud del intervalo. Véase por ejemplo la respuesta dada por ILB.

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

CIPR2. *Relación entre tamaño muestral y amplitud del intervalo.* Partiendo de la expresión del error de estimación, $Z_{\alpha/2} \times \sigma / \sqrt{n}$, se puede establecer, además de la relación descrita en la categoría anterior (CIPR1), que un aumento del tamaño de la muestra (manteniendo constante el nivel de confianza) implica que el error de estimación disminuirá (Behar, 2001). Si el objetivo es conseguir una mayor precisión en las estimaciones realizadas por medio de los intervalos de confianza, lo ideal es aumentar el tamaño muestral, ya que habitualmente no es recomendable utilizar un nivel de confianza inferior al 90%. Esta propiedad queda reflejada en la respuesta dada por JLS, donde el participante relaciona un tamaño muestral elevado con una mejor determinación del valor del parámetro poblacional bajo estudio.

La longitud del intervalo viene determinada por la muestra que hemos tomado. Si el valor de n aumentara tendríamos más información y por tanto podremos ajustar más el intervalo, es decir, disminuye. [...]. Intuitivamente, si queremos estar más seguros de dar con una franja de valores que incluya al valor real de la media, entonces cuanto más grande sea el intervalo más probabilidad hay de que englobemos en este el verdadero valor de la media (Respuesta de JLS).

CIPR3. *No se sabe si el intervalo cubre el valor del parámetro.* Cuando se señala que la obtención de los extremos del intervalo de confianza supone el desconocimiento sobre si el verdadero valor del parámetro está o no en el intervalo obtenido, como muestra la respuesta de ACG. Esta categoría está estrechamente relacionada con la interpretación correcta del nivel de confianza, dado que no se tiene la certeza que, de 100 intervalos construidos, ese intervalo sea uno de los 95 que contiene el verdadero valor del parámetro.

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

CIPR4. *Sirve para realizar contrastes.* Se han considerado aquellos casos en los que se identifica una conexión entre los intervalos de confianza y los contrastes de hipótesis. Esta propiedad viene recogida también en las investigaciones llevadas a cabo por Cumming y Fidler (2005) donde los autores resaltan de la construcción de los intervalos de confianza, la estrecha relación existente entre estos y la hipótesis nula del

contraste de hipótesis. Ejemplo de ello es la respuesta dada por BCM.

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

CIPR5. *Carácter aleatorio de los extremos*. Para cualquier muestra aleatoria de tamaño n , extraída aleatoriamente de una población normal con varianza conocida, se verifica que la media muestral \bar{X} es una variable aleatoria. Este hecho provoca que los extremos del intervalo variarán según lo haga la media muestral puesto que estos dependen de dicha variable, aun cuando la cantidad que se suma a la misma sea idéntica para diferentes muestras del mismo tamaño (Behar, 2001; Olivo, 2008). Este hecho pone de manifiesto el carácter aleatorio del intervalo, propiedad recogida en la respuesta de MVC.

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

En la Tabla 4.6.9 se ha recopilado la información frecuencial y porcentual (sobre el número total de participantes) de cada una de las propiedades que han sido mencionadas por los futuros profesores. Queremos señalar que, al igual que hay participantes que no describen ninguna propiedad fundamental del intervalo de confianza, otros de ellos indican más de una. De los 73 participantes solo 23 han nombrado al menos una de las propiedades categorizadas en este apartado, concretamente 8 han hecho referencia a una propiedad, 11 a dos propiedades y 4 a tres propiedades, por lo que, el número total asciende a 42.

Tabla 4.6.9.

Propiedades del Intervalo de Confianza Señaladas por los Participantes

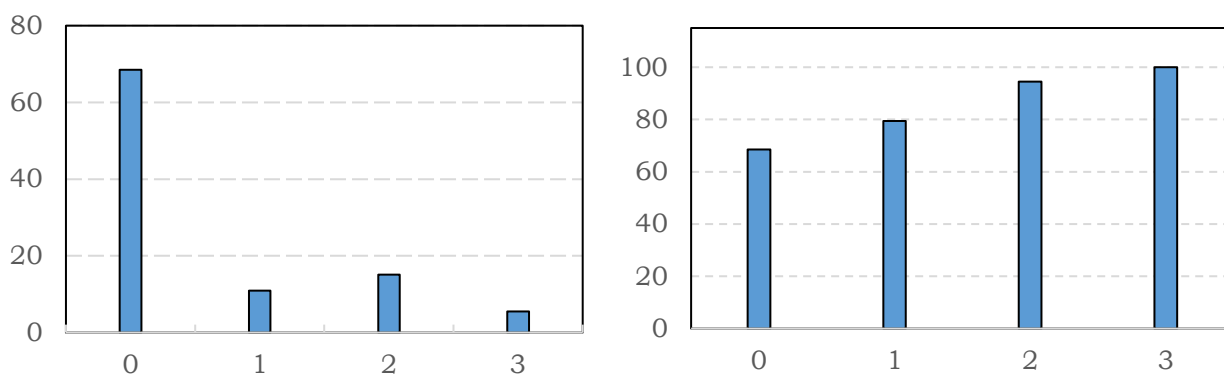
Código	Respuesta	Frecuencia	Porcentaje
CIPR1	Relación entre nivel de confianza y amplitud del intervalo	14	19,2
CIPR2	Relación entre el tamaño muestral y la amplitud del intervalo	13	17,8
CIPR3	No se sabe si el intervalo cubre el valor del parámetro	5	6,8
CIPR4	Sirve para realizar contrastes	2	2,7
CIPR5	Carácter aleatorio de los extremos	8	11,0
Total		42	57,5

Se observa que la relación entre la amplitud del intervalo, el nivel de significación

y el tamaño de la muestra seleccionada son las más destacadas ya que aproximadamente un quinto de los participantes las menciona. Le sigue el carácter aleatorio que tienen los extremos del intervalo cuando se cambia de una a otra muestra. En menor medida se ha hecho referencia a si el intervalo cubre o no el valor del parámetro y a la relación existente entre los intervalos de confianza y los contrastes de hipótesis. En este apartado no se han observado conflictos semióticos de los futuros profesores.

La distribución del número de propiedades que cita cada participante indica que lo más frecuente es no citar ninguna propiedad (Figura 4.6.2), aunque alguno de los participantes cita hasta tres. Asimismo, se observa en la gráfica de la frecuencia acumulada que son pocos los que citan más de una propiedad. Ello implica un pobre conocimiento matemático de las propiedades del intervalo de confianza por parte de los participantes

Figura 4.6.2. *Distribución y distribución acumulada (en porcentaje) del número de propiedades citadas por participante sobre intervalo de confianza*



4.6.6. Síntesis de conocimientos sobre el intervalo de confianza

Considerando el número de apartados que tienen correctos cada participante, se analiza a continuación el nivel de conocimiento que tienen los futuros profesores en relación a la elaboración del ítem sobre intervalo de confianza. Recordemos que una correcta elaboración del mismo supone describir correctamente el planteamiento del problema para posteriormente realizar los cálculos necesarios con objeto de obtener los valores

que determinan los extremos del intervalo. Además, es necesario que el participante indique una definición e interpretación correcta de los resultados del mismo. Así pues, una puntuación de 4 puntos supone una correcta elaboración del ítem.

En la Tabla 4.6.10 se ha recogido la información sobre el número de apartados correctos dado por cada participante en el problema sobre intervalo de confianza. Comparando la información recogida en la Tabla 4.6.10 con la realizada para el contraste de hipótesis (véase Tabla 4.5.11) se observa que en este caso seis han respondido adecuadamente a los cuatro apartados evaluados. Sin embargos se ha detectado mayores problemas a la hora de dar respuesta a este ítem ya que hay 9 participantes no han realizado correctamente algún apartado. Este resultado es de especial interés ya que, aunque son varios los conceptos a dominar para la ejecución de este tipo de tareas, muchos de los estudiantes reconocen y elaboran un procedimiento puramente mecánico con objeto de obtener los extremos del intervalo.

Tabla 4.6.10.

Distribución del Número de Apartados Correctos Sobre el Intervalo de Confianza

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

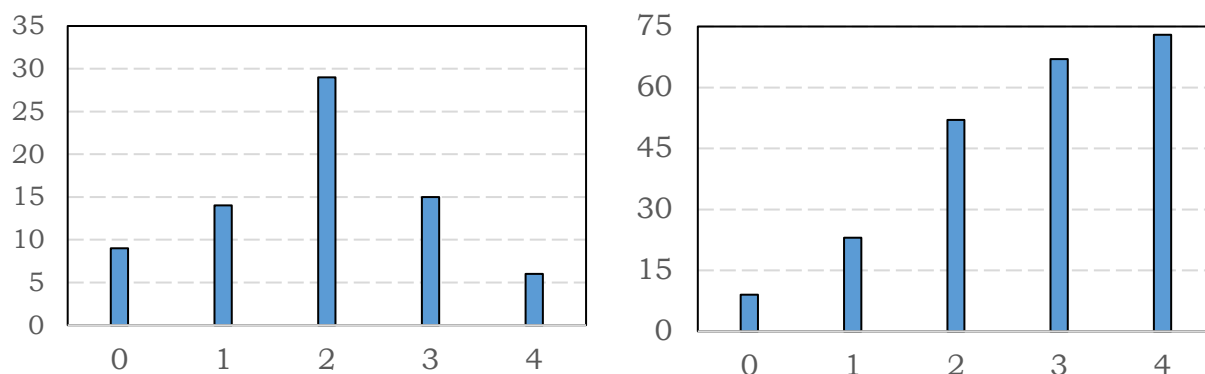
Un análisis detallado de los apartados clasificados correctamente nos ha permitido ver que los participantes con uno o dos apartados correctos han realizado un planteamiento adecuado del ítem, completándose, en su mayoría, para el caso de dos apartados correctos, con la obtención de los valores numéricos de los extremos del intervalo solicitado. Se han detectado mayores dificultades, por parte de los futuros profesores, a la hora de definir qué es un intervalo de confianza, pues de los 15 que contestan correctamente tres apartados, 10 de ellos no llegan a definir tal término.

La representación gráfica de la distribución del número de apartados correctos es casi simétrica respecto al valor 2, pues deja por encima y por debajo de él aproximadamente el 30% de la distribución (véase Figura 4.6.3). Observamos además

un salto significativo entre 1 y 2 apartados correctos pues se produce un incremento de aproximadamente un 40%, dicho crecimiento se ve más suavizado a medida que aumentan el número de apartados correctos.

Las conclusiones del análisis revelan las dificultades que presentan los futuros profesores cuando abordan problemas de construcción, interpretación y definición de intervalos de confianza. Este hecho, junto a los resultados obtenidos cuando se le ha preguntado sobre el número de propiedades fundamentales que recordaban (véase Tabla 4.6.9), pone de manifiesto un conocimiento débil sobre la determinación de intervalos de confianza. Se recuerda que el contenido de intervalos de confianza sigue manteniéndose en el actual currículo y continúa siendo uno de los problemas que forman parte en las pruebas de acceso a la universidad actuales. Por tanto, se debe trabajar en la dirección correcta para que los futuros profesores adquieran un mejor conocimiento sobre el tema con el fin asegurar un proceso de enseñanza y aprendizaje de calidad.

Figura 4.6.3. *Distribución y Distribución Acumulada del Número de Apartados Correctos por Participante en el Ítem sobre Intervalos de Confianza*



Para finalizar la síntesis, en la Tabla 4.6.11 se recopilan los conflictos semióticos que sobre el intervalo de confianza se han determinado en las respuestas de los estudiantes. Aunque aparecen bastantes conflictos, son menos que los que se encontraron sobre el contraste de hipótesis, debido a que es menor el número de objetos matemáticos implicados en este tema en comparación con el primero. Por tanto, se ha de dar la razón a estudios como los de Cumming y Fidler (2005) o Cumming, William y Fidler (2004) que indican que los intervalos de confianza tienen una interpretación más

sencilla que los contrastes. Esto justifica posiblemente que se hayan suprimido los contrastes de hipótesis en el Bachillerato en las últimas orientaciones curriculares.

También en este caso es mayor el número de conflictos semióticos conceptuales que procedimentales, lo que indica que la enseñanza se concentra más en el aprendizaje de fórmulas y no se logra una comprensión profunda por parte de los estudiantes.

Tabla 4.6.11.

Lista de Conflictos Semióticos Identificados en la Resolución de la Tarea Sobre Intervalo de Confianza

<i>Conceptuales</i>
1. Confundir estadístico (media muestral) y parámetro. (media poblacional) (CIP3, CIC5)
2. Definir el coeficiente de confianza como probabilidad de acierto en la estimación o su complementario como probabilidad de error en la estimación (CIP2)
3. Definir el coeficiente de confianza como la probabilidad de que el parámetro esté comprendido en el mismos (CII3)
4. Pensar que el intervalo de confianza contiene a la media muestral (CII4)
5. Visualizar los extremos del intervalo como constantes (CII3)
6. Concepción determinista del intervalo de confianza (CIP3, CII2)
7. Describir el intervalo como “varios pares de números” (CIP2)
8. Pensar que es posible comprobar si el intervalo contiene o no a la media de la población (CIP3)
9. Interpretación bayesiana del intervalo (CID2)
<i>Procedimentales</i>
1. No recordar la expresión de cálculo del intervalo de confianza (CIP2)
2. Error en la fórmula de cálculo de los extremos del intervalo (CIP4)
3. Uso incorrecto u olvido de la tabla normal (CIC2, CIC4)
4. Errores de cálculo (CIC3)
5. Dificultad en el trabajo con desigualdades (CIC2)
6. Dificultad en identificar los datos del enunciado (CIC4)

4.6.7. Conclusiones sobre el conocimiento común del intervalo de confianza

Los resultados obtenidos sobre el intervalo de confianza muestran que la mayoría de los participantes realizan un planteamiento adecuado y un cálculo correcto de los extremos del intervalo. Sin embargo, estos buenos resultados sufren un cambio drástico cuando se analiza la información referente a la interpretación y definición del intervalo de confianza donde el porcentaje de respuesta sin interpretación es del 41,1% y sin definición del 64,4%. La mayor parte de las propiedades se han centrado en destacar la relación que existe entre el ancho del intervalo con el tamaño muestral y el coeficiente de confianza y en el carácter aleatorio que tienen los extremos del intervalo. Sin embargo, solamente un tercio ha reconocido alguna propiedad de los intervalos.

Algunos de los errores detectados se han identificado en investigaciones previas, Behar (2001), Cumming y Fidler (2005), Garfield et al. (2004), Harradine et al. (2011), Olivo (2008) y Vallecillos (1999), entre otros. Dado que el tema en estudio forma parte de la educación preuniversitaria y de muchos grados universitarios en España, es necesario asegurar principalmente que los futuros profesores posean un adecuado conocimiento sobre el tema ya que ellos son los encargados de transmitirlo en el aula. Una correcta y clara transmisión, profundizando tanto en la interpretación como en la definición del mismo, puede garantizar un aprendizaje más significativo por parte de los estudiantes.

Asimismo, la importancia que tiene el uso de los intervalos de confianza en las investigaciones también ha sido resaltado en el manual de la asociación profesional American Psychological Association (APA) recomendado hacer uso de la información que aporta los intervalos de confianza con objeto de tener un “mejor informe” (APA, 2010, p. 34).

4.7. CONOCIMIENTOS DIDÁCTICOS SOBRE INFERENCIA ESTADÍSTICA

El tercer objetivo de este estudio fue evaluar algunos elementos de la faceta cognitiva del conocimiento didáctico matemático de los futuros profesores sobre contrastes de hipótesis e intervalos de confianza.

Para cumplir con este objetivo, se analizan en esta sección las respuestas dadas a los ítems 3 y 4 (véase subsección 4.4.2) propuestos a los participantes. En dichos ítems se pregunta por los posibles errores que podrían cometer sus estudiantes al realizar problemas sobre contrastes de hipótesis e intervalos de confianza. Usaremos el término error para analizar sus respuestas, pues a los futuros profesores no se les introdujo en la idea de conflicto semiótico, aunque, en nuestro marco teórico, la mayor parte de estos errores citados por los futuros profesores se pueden interpretar en términos de conflicto semiótico.

La elaboración de dichos ítems se llevó a cabo en la segunda sesión del taller. De los 73 participantes iniciales solamente 70 realizaron las actividades enfocadas en evaluar su conocimiento didáctico en el tema. En total se analizaron 491 errores, lo que corresponde, en promedio, a 7,01 errores por cada futuro profesor.

La Tabla 4.7.1 resume la frecuencia y porcentaje de errores citados en cada apartado del análisis y, la Tabla 4.7.2 muestra el número y porcentaje de estudiantes que citan al menos un error en cada apartado. Se observa un buen desempeño en los futuros profesores y un mayor número de errores citados respecto al contraste de hipótesis, procedimiento más laborioso y que involucrar mayor número de conceptos y donde vimos que ellos mismos presentan un mayor número de conflictos semióticos.

Los errores más citados en ambas tareas son los procedimentales, seguidos por los de interpretación de resultados mientras que en su propio desempeño fueron más frecuentes los conceptuales, lo que indica que los futuros profesores no son totalmente conscientes de los conflictos semióticos que se comentaron en las anteriores secciones. En el contraste de hipótesis se citan a continuación: el planteamiento de las hipótesis y conceptuales y, en el intervalo los errores de tipo conceptual. Hay poca consciencia de los errores de planteamiento del procedimiento en ambas tareas.

Tabla 4.7.1.

Número de Errores Citados y Porcentaje (respecto al total de errores)

Categorías	Contraste		Intervalo	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Planteamiento de las hipótesis	55	18,6	--	--
Planteamiento del procedimiento	28	9,5	11	5,6
Errores conceptuales	53	17,9	28	14,4
Errores procedimentales	90	30,4	95	48,7
Interpretación de resultados	70	23,6	61	31,3
Total	296	100	195	100

En la Tabla 4.7.2 se describe la frecuencia de errores que cita cada profesor en las diferentes categorías. Se observa que en todas las categorías hay profesores que no citan errores y que son más los que citan en el contraste que en intervalo, y de nuevo más procedimentales que conceptuales.

Seguidamente se clasifican los errores citados, teniendo en cuenta los identificados en las investigaciones previas, mostrando ejemplos de las respuestas de los futuros profesores. Puesto que algunos participantes citan más de un error en el mismo apartado, se ha realizado el cálculo de los porcentajes en función al número total de participantes.

Tabla 4.7.2.

Número de Futuros Profesores que Citan Algún Error en cada Categoría y Porcentaje (respecto al total de futuros profesores)

Categorías	Contraste		Intervalo	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Planteamiento de las hipótesis	46	65,7	--	--
Planteamiento del procedimiento	24	34,3	9	12,9
Errores conceptuales	38	54,3	26	37,1
Errores procedimentales	51	72,9	51	72,9
Interpretación de resultados	50	71,4	49	70,0

4.7.1. Errores en el planteamiento de las hipótesis

Un total de 55 errores se citaron en esta categoría, que es exclusiva del contraste de hipótesis y que pueden clasificarse en la forma siguiente:

CEH1. *Confundir hipótesis nula y alternativa.* En esta categoría se incluyen las respuestas que señalan la posible confusión entre las hipótesis nula y alternativa, error descrito por Batanero (2000), Díaz-Batanero (2018) y Vallecillos (1999) y que apareció en algunos estudiantes en el planteamiento de hipótesis en el ítem 1. A continuación, se muestra un ejemplo.

Considero que el mayor problema que se le presenta al alumnado en los contrastes de hipótesis es saber diferenciar entre la hipótesis nula y la hipótesis alternativa, cometiendo el error al confundirlas y por lo tanto teniendo un contraste erróneo (Respuesta de MGG).

CEH2. *Plantear hipótesis no complementarias.* Otros futuros profesores aluden a la posibilidad de construir hipótesis no complementarias, error recogido en investigaciones realizadas con estudiantes (Cañadas et al., 2016; Espinel et al., 2007; Vallecillos, 1999). En dicho caso las hipótesis del contraste no cubrirían en su totalidad los valores posibles del espacio paramétrico, por lo que el resultado podría no ser concluyente. Como ejemplo se muestra la respuesta de JLS.

A la hora de hacer un contraste de hipótesis, si por ejemplo tenemos una distribución normal, los estudiantes podrían no plantear correctamente las hipótesis nula o alternativa (deben de ser complementarias) (Respuesta de JLS).

CEH3. *Respuesta imprecisa.* Junto a los errores manifestados anteriormente, algunos estudiantes dan respuestas imprecisas, como la siguiente en la que se sugiere que puede haber un error al plantear las hipótesis, pero no se indica en qué consiste el

error.

No saber identificar el contraste, por ejemplo, dar una hipótesis nula o alternativa errónea (Respuesta de MLC).

En la Tabla 4.7.3 se muestra el desglose de los 55 errores asociados al planteamiento de la hipótesis y el porcentaje de estudiantes que cita cada uno de ellos. La mayor parte de los errores citados se relacionan con la confusión entre la hipótesis nula y alternativa. Es menos frecuente reconocer que las hipótesis podrían no ser complementarias y un 20% cita otros errores o son imprecisos al dar sus respuestas.

Tabla 4.7.3.

Frecuencia de Errores en el Planteamiento de las Hipótesis y Porcentaje de Estudiantes que lo Citan

Código	Respuesta	Frecuencia	Porcentaje
CEH1	Confundir hipótesis nula y alternativa	35	50,0
CEH2	Plantear hipótesis no complementarias	6	8,6
CEH3	Respuesta imprecisa	14	20,0

Por otro lado, se ha observado que no se han descrito los errores consistentes en hacer uso de los estadísticos muestrales en el planteamiento de las hipótesis, confundir una hipótesis unilateral con otra bilateral, o uso de una notación no adecuada descritos en diferentes investigaciones (Cañadas et al., 2016; Espinel et al., 2007; Harradine et al., 2011; Ramos et al., 2009; Vallecillos, 1999). Se resalta el hecho de que algunos conflictos identificados en la muestra al plantear las hipótesis (como olvidar plantear una de las hipótesis o no comprender que la hipótesis nula ha de ser puntual) no fueron citados, por lo que los profesores que los manifestaron no son conscientes de ellos. Así mismo, se tiene que un 30% de los futuros profesores no describen ningún error en el planteamiento de las hipótesis, lo que sugiere que no son conscientes de la importancia de este tema dentro del procedimiento.

4.7.2. Errores en la selección del contraste de hipótesis e intervalo de confianza

Se analiza los errores que se han descrito en referencia con la correcta selección de un contraste de hipótesis o un intervalo de confianza adecuado para resolver los problemas planteados, que han sido los siguientes:

CES1. *Error al elegir la distribución muestral.* La distribución muestral adecuada es fundamental, tanto para realizar un contraste, como para calcular un intervalo de confianza, debido a que todos los cálculos posteriores, por ejemplo, la determinación del p -valor, o el cálculo de la amplitud del intervalo de confianza se basan en ella.

El error en la elección de la distribución muestral inadecuada, por ejemplo, utilizar la distribución t en lugar de la normal o tomar unos parámetros erróneos para la distribución normal que sirve de modelo a la distribución muestral del estadístico es señalado por algunos estudiantes y fue encontrado en la investigación de Espinel et al. (2007). Como ejemplo se muestra la respuesta de IPM, que indica varias posibilidades que llevan a una distribución muestral incorrecta, como no aplicar el teorema central del límite o calcular mal los parámetros de la distribución muestral. Incluimos también el error descrito por ACG que consiste en expresar el intervalo en función de la distribución normal $N(0,1)$. Se recuerda que algunos estudiantes tomaron erróneamente la distribución t para sus cálculos en los problemas propuestos y otros olvidaron o confundieron los parámetros de la distribución muestral.

Al aplicar el Teorema Central del Límite, puede que algunos estudiantes no lo apliquen y no sepan qué distribución sigue, lo que hace que no lleguen a plantear el intervalo de confianza. Otro posible error relacionado con el Teorema Central del Límite es que lo apliquen mal o que no dividan entre \sqrt{n} , o incluso que dividan por n y no por su raíz, lo que provoca un error grave en los cálculos (Respuesta de IPM).

[...] una vez tipificado, el alumno puede olvidar deshacer el cambio y presentar el resultado final con la variable tipificada (Respuesta de ACG).

CES2. *No identificación del tipo de contraste o del tipo de intervalo.* Estas respuestas describen los errores que pueden cometer los estudiantes cuando fallan en la elección del tipo de contraste o intervalo (unilateral o bilateral), con lo cual cometerán errores posteriores. Así, si el contraste es bilateral, se deben construir dos regiones (a izquierda y derecha con probabilidad $\alpha/2$) mientras que si es unilateral únicamente se contará con una región de rechazo, a izquierda o derecha según el tipo de contraste, con probabilidad de α .

El mismo problema se presenta con la diferencia entre intervalo de confianza unilateral o bilateral. A continuación, se muestra la respuesta dada por JCR, que alude a este error, que fue descrito por Espinel et al. (2007). Recordamos que fueron bastantes los futuros profesores que en el contraste de hipótesis confundieron el contraste

unilateral con bilateral o el unilateral izquierdo con el derecho.

Confusiones con el carácter bilateral/unilateral de las colas de distribuciones (Respuesta de JRC).

CES3. *Respuesta imprecisa.* Se han considerado aquellas respuestas en las que el estudiante no indica claramente en qué consiste el error de planteamiento, lo que podría enmascarar un desconocimiento de estos errores. Ejemplo de ello es la respuesta dada por MTF en la que no se detecta el posible error que puede cometer el alumno para no identificar de una forma correcta el tipo de contraste a realizar.

No identificar el contraste adecuadamente (Respuesta de MTF)

De la información recogida en la Tabla 4.7.4 se observa que fueron pocos los participantes que han señalado posibles errores en la elección del contraste de hipótesis o del intervalo de confianza adecuado, y la mayoría de los que los citan en el contraste de hipótesis son imprecisos, lo que indica que este tipo de errores, posibles de cometer al seleccionar el contraste o intervalo adecuado y con repercusión sobre los resultados, es poco conocido por los estudiantes de la muestra.

Tabla 4.7.4.

Frecuencia de Errores en el Planteamiento (N.) y Porcentaje (%) de Participantes que lo Citan

Código	Respuesta	Contraste		Intervalo		Aparece en la muestra
		N.	%	N.	%	
CES1	Error en la elección de la distribución muestral	8	11,4	5	7,1	Si
CES2	No identificación del tipo de contraste (bilateral o unilateral)	8	11,4	4	5,7	Si
CES3	Respuesta imprecisa	12	17,1	2	2,9	

4.7.3. Errores conceptuales

Son muchos los conceptos que intervienen en la formulación, tanto del contraste de hipótesis, como del intervalo de confianza y algunos son comunes a los dos procedimientos, aunque otros son específicos. En la evaluación del conocimiento matemático respecto a los dos procedimientos, se identificaron un gran número de

conflictos semióticos por parte de los futuros profesores de la muestra. Es interesante conocer si, una vez corregidos los problemas se han hecho conscientes de su existencia y los prevén en sus futuros estudiantes.

Respecto a los conceptos comunes a los dos problemas, los participantes en el estudio citaron los siguientes errores:

CEC1. *Confundir estadístico y parámetro.* En esta categoría se han considerado aquellas respuestas en las que se confunden los parámetros poblacionales (resúmenes desconocidos y constantes de una población) con los estadísticos muestrales (valores conocidos y que varían de una muestra a otra) (Vallecillos, 1999). Esta confusión está relacionada con la falta de discriminación de algunos estudiantes del Máster entre los tres tipos de distribución que intervienen en inferencia: distribución de los datos de la muestra, distribución de la variable en la población y distribución muestral (Harradine et al., 2011; Schuyten, 2011). Algunos participantes en el estudio citaron esta confusión, que varios de ellos cometieron, tanto al plantear sus hipótesis, como en la construcción del intervalo de confianza:

Para empezar, podrían confundir media muestral y media de la distribución, o en el caso de los contrastes en variables que siguen distribuciones binomiales, la proporción muestral con la probabilidad de éxito de la distribución (Respuesta de DCC).

CEC2. *Aplicar erróneamente el teorema central del límite.* Este teorema asegura que la suma de n variables aleatorias independientes provenientes de una población (no necesariamente normales) con varianza finita, tiende a distribuirse normalmente a medida que aumenta el tamaño muestral y permite identificar la distribución de la media muestral. Para poder aplicarlo, en primer lugar, es necesario que el tamaño de la muestra sea adecuado y el muestreo sea aleatorio. Además, se deben asignar unos valores concretos a la media y desviación típica de la distribución normal resultante. Algunos participantes como IPM indican que un error podría ser su aplicación incorrecta, debido a utilizar unos parámetros erróneos o incluso a no aplicarlo.

Al formar la distribución de la media muestral, es posible que no aplique bien el Teorema Central del Límite, se olvide de dividir por \sqrt{n} o sólo divida por n . Otro posible error es que directamente no aplique el Teorema Central del Límite (Respuesta de IPM).

CEC3. *Interpretar incorrectamente el nivel de significación o el nivel de confianza.* Algunos futuros profesores sugieren errores de comprensión del nivel de significación,

aunque generalmente son poco precisos, pues no se indica con claridad si el error consiste en el cambio entre los dos términos de la probabilidad condicionada o en la creencia de que el nivel de significación es la probabilidad de equivocarse al rechazar la hipótesis nula, que son algunos de los errores asociados a dicho concepto (Birnbaum, 1982; Falk, 1986, Vallecillos, 1999).

Por tanto, no queda claro si realmente los participantes conocen los errores típicos asociados a la interpretación del nivel de significación. Como ejemplo, se muestra la respuesta dada por MEG. Por otro lado, AVE sugiere un error no citado en las investigaciones previas, que consiste en confundir el nivel de significación con su complementario (nivel de confianza).

Comprender el concepto de nivel de significación (Respuesta de MEG).
Confundir el nivel de significación con el nivel de confianza (Respuesta de AVE).

CEC4. *Confundir otros conceptos.* Algunos participantes sugieren la confusión entre conceptos propios de estadística descriptiva (como confundir desviación típica y varianza) o relacionados con la estimación puntual y no con el contraste de hipótesis o el intervalo de confianza; por ejemplo, confundir el estimador de la proporción con su complementario. Este conflicto no se encontró en la muestra de futuros profesores.

Confundir la varianza y la desviación típica. (Respuesta de MGG).
En el problema: “Tomada, al azar, una muestra de 120 estudiantes de una Universidad se encontró que 66 de ellos no hablaban inglés,.”, el alumno suele tomar $\hat{p} = 66/120$ en lugar de $\hat{p} = 140/120$ (Respuesta de MVC).

Por otro lado, se citaron los siguientes errores propios únicamente del contraste de hipótesis;

CEC5. *Interpretar incorrectamente el p-valor.* El *p*-valor es definido como la probabilidad de obtener el valor dado u otro más extremo bajo la hipótesis nula. Sin embargo, dicho término es interpretado por algunos estudiantes como la probabilidad de que la hipótesis sea cierta si se obtuvo el valor dado del estadístico o bien se interpreta como probabilidad de que el valor obtenido del estadístico se deba al azar o como probabilidad de la hipótesis nula (Batanero, 2000). Ninguno de estos errores fue citado por los participantes, aunque varios los mostraron durante la resolución del problema planteado, pues sólo MSM hace referencia al mismo e indica que se confunde con el nivel de significación. Esta confusión es importante, pues mientras el nivel de

significación es constante y lo fija el investigador antes de realizar el contraste, el p -valor es variable y se calcula como consecuencia del mismo.

Confusión del nivel de significación y el p -valor (Respuesta de MSM).

CEC6. *Confundir los errores Tipo I y Tipo II.* Una vez tomada la decisión sobre la hipótesis nula, existe la posibilidad de cometer errores que dependen de la decisión tomada: Error Tipo I es el que se comete cuando la hipótesis nula es rechazada siendo cierta y el Error Tipo II cuando la hipótesis nula no es rechazada siendo falsa. La importancia de diferenciarlos se debe a que la probabilidad asociada al primer error (nivel de significación) queda determinada antes de la realización del mismo, mientras que la probabilidad asociada al segundo error es variable y depende de varios factores. Este tipo de conflicto no apareció en nuestra muestra, al no preguntarles por los tipos de error en un contraste. Un ejemplo de cita de este error se reproduce a continuación:

No distinguir los errores de tipo I y los de tipo II (Respuesta de EHB).

El análisis de las respuestas nos ha permitido observar que más de la mitad de los participantes (53) describieron algún error conceptual. De la información de la Tabla 4.7.5 se tiene que el error más citado es la confusión entre estadístico y parámetro el 71,5% de los mismos.

Tabla 4.7.5.

Frecuencia (N) y Porcentaje (%) de Estudiantes que Cita cada Error Conceptual

Código	Respuesta	Contraste		Intervalo		Aparece en la muestra
		N.	%	N.	%	
CEC1	Confundir estadístico y parámetro	30	42,9	20	28,6	Sí
CEC2	Aplicar incorrectamente el teorema central del límite	4	5,7			Sí
CEC3	Interpretar incorrectamente el nivel de significación o el nivel de confianza	6	8,6	4	5,7	Sí
CEC4	Confundir otros conceptos			4	5,7	No
CEC5	Interpretar incorrectamente el p -valor	2	2,9			Sí
CEC6	Confundir los errores Tipo I y Tipo II	10	14,3			No

El segundo error conceptual más citado es la confusión entre los tipos de errores en un contraste de hipótesis. Apenas se ha aludido a los errores de interpretación del

nivel de significación y del p -valor y cuando se hace es en forma muy imprecisa. Ello nos hace sospechar que los participantes pudieran no ser conscientes de los errores relacionados con la interpretación de estos conceptos que, de acuerdo a la investigación previa, están muy extendidos. Otros errores conceptuales son citados con poca frecuencia. El número de citas de errores conceptuales asociados al contraste de hipótesis duplica a los relacionados con los intervalos de confianza, lo que sugiere que los participantes encuentran conceptualmente más difícil el primer procedimiento y le asocian más conceptos diferentes.

En resumen, los futuros profesores fueron capaces de citar algunos de los principales errores conceptuales que ellos mismos mostraron al resolver los problemas. No se citan los relacionados con el planteamiento de las hipótesis o la confusión de un contraste unilateral y bilateral, las condiciones en que se puede modelizar la distribución muestral mediante la distribución normal o la concesión determinista del contraste de hipótesis. Respecto al intervalo no se indica la posibilidad de pensar que el intervalo se refiere a los valores de la variable o la media muestral. En definitiva, aunque la actividad formativa permitió alertar a los futuros profesores de los posibles conflictos semióticos conceptuales de sus estudiantes, quedaron muchos sin asimilar. Sería, entonces necesario ampliar la acción formativa descrita en este trabajo.

4.7.4. Errores procedimentales

Una vez seleccionado el contraste o intervalo de confianza adecuado se requiere una serie de pasos para completarlo, en los cuales pueden producirse diferentes errores. Los participantes del estudio han citado los siguientes diez errores:

CEP1. *Error en la identificación de los datos.* El primer paso del procedimiento consiste en identificar los diferentes datos del enunciado, por ejemplo, tamaño de la muestra, valor de la media muestral o tipo de distribución en la población. En esta identificación pueden cometerse errores, como sucedió en los participantes de la muestra y como citan algunos de ellos:

Interpretar los datos que nos da el problema (Respuesta de LCM).

CEP2. *Determinación del estadístico de contraste.* La realización de un contraste o la construcción de un intervalo requieren la elección de un estadístico en la muestra

que esté relacionado con el parámetro que se desea estimar o contrastar. Por ejemplo, para un intervalo de confianza de la media poblacional se debe utilizar la media muestral por ser un estimador insesgado de la misma. Un error señalado por algunos participantes es la elección inadecuada del estadístico, que no se presentó en nuestra muestra.

Confundir el estadístico que hay que calcular para comprobar con el valor teórico. (Respuesta de JRC).

CEP3. *Error de tipificación o de operación con las desigualdades.* Generalmente la distribución muestral es normal, con una media y varianza que se determinan a partir de las características de la población en estudio y el tamaño de la muestra. Para el cálculo del p -valor, regiones o puntos críticos, se recurre a la tipificación de la variable, para convertirla en una normal estándar. Ello requiere una serie de operaciones con desigualdades donde el estudiante podría cometer errores, como lo recoge la respuesta de IPM para el caso del intervalo de confianza. Ambos errores se presentaron en la muestra del estudio:

Como los estudiantes trabajan con la distribución $N(0,1)$, deben tipificar, y es posible que no lo hagan, lo que causaría un grave error puesto que no saben calcular los datos en la distribución normal diferente, o también es posible que tipifiquen de forma errónea (Respuesta de IPM).

CEP4. *Error en el cálculo de probabilidades.* En algunos casos se citan los errores de cálculo de probabilidades, como, por ejemplo, calcular probabilidades dentro de un intervalo, que requiere hallar la diferencia entre las probabilidades de que la variable tome un valor menor o igual que cada uno de los extremos del intervalo. Algunos participantes lo citan, aunque de hecho no apareció en la evaluación de sus propios conocimientos:

Cómo operar con las probabilidades por el extremo superior e inferior del intervalo $P(-a < Z \leq a) = P(Z \leq a) - P(Z > a) = P(Z \leq a) - [1 - P(Z \leq a)]$ (Respuesta de EJO).

CEP5. *Lectura incorrecta de las tablas estadísticas.* Independientemente del proceso empleado, el uso adecuado de las tablas estadísticas adquiere cierta importancia pues un error en la lectura de la tabla podría suponer la determinación de regiones de rechazo o amplitudes de intervalos incorrectas, entre otros. La lectura de las tablas tiene su dificultad, descrita por Espinel et al. (2007), debido a la diversidad

de tablas existentes, lo que ha sido señalado por los participantes. Se recuerda que varios participantes en el estudio fueron ellos mismos incapaces de leer correctamente las tablas de la distribución normal:

Es normal que se equivoquen en la lectura de las tablas (Respuesta de LPM).

CEP6. *Error en el cálculo del p-valor.* El cálculo del *p*-valor puede ser erróneo cuando se calcula como una probabilidad simple, en lugar de condicionada, error descrito por Falk (1986) y cometido en la muestra participante, o bien si, en lugar de tomar la desigualdad apropiada para el tipo de contraste planteada (por ejemplo, la probabilidad que corresponde a los valores que superan el obtenido para el estadístico) se toma la contraria. Los estudiantes que indican este error procedimental suelen ser muy imprecisos, como el siguiente:

No calcular bien el *p*-valor (Respuesta de ILG).

CEP7. *Determinación del punto crítico.* Mientras que en el cálculo del *p*-valor se parte de un valor del estadístico para calcular la probabilidad, para determinar el punto crítico se realiza la operación inversa, pues se parte de un valor de probabilidad (nivel de significación) para determinar el valor del estadístico que corresponde a dicha probabilidad, lo que es citado en forma poco concreta por algunos participantes. En el estudio este conflicto aparece sobre todo cuando confunden el contraste unilateral y bilateral:

Podrían cometer errores al calcular el valor crítico $Z_{1-\alpha}$ (Respuesta de AGB).

C8EP. *Determinación incorrecta de las regiones de rechazo/no rechazo.* La región de rechazo está constituida por aquellos valores del estadístico de contraste que se alejan mucho de la hipótesis nula, siendo poco probable obtener dichos valores si la hipótesis nula es cierta. Un error relacionado consiste en intercambiar las regiones de un contraste, descrito por Vallecillos (1994), que es citado por algunos futuros profesores y apareció en la muestra en los participantes que usaron el método de máxima verosimilitud:

Un error común puede ser el realizar tomar la región de rechazo en los dos extremos (Bilateral) en lugar de tomarla solo en uno (Unilateral) (Respuesta de AGS).

CEP9. *Errores al establecer los límites del intervalo.* La obtención del intervalo de confianza para la media poblacional implica sumar y restar a la media muestral una fórmula en la que interviene el valor crítico, el tamaño de la muestra y la desviación típica. Pueden obtenerse extremos erróneos si se confunde alguno de estos valores, como indica JSM. Fue también cometido en la muestra.

Poner el intervalo al revés y no percatarse, el mayor valor lo ponen a la izquierda y el menor a la derecha: (a,b) con $a > b$ (Respuesta de JSM).

CEP10. *Errores de cálculos o en la terminología.* Dentro de esta categoría se han incluido aquellas respuestas que hacen alusión a los posibles errores de cálculo que se cometen en la elaboración del contraste o al uso de la terminología; ambos conflictos ocurrieron en la primera parte del estudio de evaluación.

A la hora de calcular el tamaño de la muestra, si el resultado obtenido es decimal suele no redondear o redondear al valor más próximo y no por exceso como es necesario (Respuesta de MVC).

La Tabla 4.7.6 resume los resultados obtenidos del análisis realizado en relación a los errores procedimentales que citan los futuros profesores. Se observa que este tipo de errores son bastante variados y numerosos, se citan aproximadamente con la misma frecuencia en relación a los contrastes y a los intervalos y la mayor parte de las veces hay poca precisión al describirlos.

El error procedimental que parece más claro es el de lectura de las tablas de la distribución normal, tipificación, identificados por Espinel et al. (2007), y determinación de las regiones. Observamos además el incremento que se produce (más del doble) en los errores asociados a la identificación de los datos en la construcción del intervalo de confianza respecto a la elaboración del contraste de hipótesis; este resultado es algo llamativo ya que los datos empleados tanto en el contraste como en el intervalo son exactamente los mismos. El resto de errores son descritos en forma muy difusa.

La mayoría de errores corresponden a los conflictos semióticos detectados en las respuestas de los futuros profesores al resolver los problemas y el número diferente citado es mayor que en los conceptuales, aunque, de hecho, los futuros profesores mostraron más conflictos conceptuales que procedimentales. Es otra indicación de que su aprendizaje es de reglas y no son conscientes de su propia falta de comprensión profunda del tema.

Tabla 4.7.6.

Frecuencia de Estudiantes que Cita cada Error Procedimental (N.) y Porcentaje (respecto al total de futuros profesores)

Códigos	Respuestas	Contraste		Intervalo	
		N.	%	N.	%
CEP1	Error en la identificación de los datos	7	10,0	16	22,9
CEP2	Determinación del estadístico de contraste	10	14,3	3	4,3
CEP3	Error de tipificación o al operar con desigualdades	18	25,7	38	54,3
CEP4	Error en el cálculo de probabilidades	5	7,1	2	2,9
CEP5	Lectura incorrecta de las tablas estadísticas	23	32,9	21	30,0
CEP6	Error en el cálculo del p -valor	2	2,9		
CEP7	Determinación del punto crítico	6	8,6		
CEP8	Determinación incorrecta de las regiones	9	12,9		
CEP9	Error al establecer los límites del intervalo			10	14,3
CEP10	Errores de cálculo o en la terminología	10	14,3	5	7,1

4.7.5. Errores de interpretación

La fase final es la interpretación de los resultados de los cálculos matemáticos, que depende de si el procedimiento realizado es el contraste de hipótesis o el intervalo de confianza. Respecto al contraste de hipótesis se citan los siguientes errores de interpretación:

CEI1. *No saber cuándo rechazar o no rechazar la hipótesis nula.* Por la experiencia docente en clases particulares o de otro tipo, se indica que el estudiante sabe aplicar el procedimiento, pero no llega a tomar la decisión, pues no comprende a fondo el proceso por el cual se toma la decisión. Este es un error poco citado en la literatura, pero algunos futuros profesores no llegaron a tomar una decisión; de ahí que citen este error.

También cometía errores al ver si rechazaba o no la hipótesis, no era capaz de retener cuándo se rechazaba o cuándo no (Respuesta de ACR).

CEI2. *Utilización del término “acepta” la hipótesis nula.* El contraste de hipótesis está basado en la información proporcionada por la muestra, por lo que, si la hipótesis nula es rechazada, sólo se puede indicar que los datos de la muestra ofrecen cierta evidencia sobre su falsedad (Batanero, 2000). Por el contrario, si los datos contradicen la hipótesis, esta se rechaza. Algunos futuros profesores son conscientes de que los estudiantes podrían utilizar el término aceptación, que no es conveniente. No aparece en la muestra del estudio:

Que, si les sale que hay que aceptar el contraste, es decir, si el p -valor es mayor que el

valor α , ellos digan que se acepta la hipótesis nula en vez de decir que no hay evidencias para rechazar la hipótesis nula (Respuesta de BDM).

CEI3. *Cometer un error tipo I o un error tipo II.* Es decir, aceptar la hipótesis nula siendo falsa o bien rechazar la hipótesis nula siendo verdadera. Estos son los posibles errores que se pueden cometer al realizar un contraste de hipótesis, incluso cuando el procedimiento se lleve a cabo correctamente, pues son debidos a la aleatoriedad del proceso de muestreo que, aunque se pueda disminuir, nunca desaparece por completo. Los futuros profesores han aludido a este tipo de error en algunas de sus respuestas, como ILG que indica los dos errores.

Cometer errores del Tipo I o II al realizar la interpretación, es decir, rechazar cuando la hipótesis es cierta o no rechazar cuando la hipótesis es falsa (Respuesta de ILG).

CEI4. *No contextualizar el resultado.* Algunos estudiantes realizan todos los cálculos matemáticos llegando a tomar una decisión correcta en relación al rechazo o no de la hipótesis nula, sin embargo, no interpretan dicho resultado dentro del contexto del problema. Esta falta de competencia de contextualización ha sido descrita en el siguiente ejemplo.

No saber extraer la conclusión pertinente (Respuesta de MTF).

En relación a los posibles errores de interpretación del intervalo de confianza, hemos encontrado los siguientes:

CEI5. *Interpretación del nivel de confianza o nivel de significación.* Dependiendo del libro de texto utilizado se da como dato en la construcción del intervalo de confianza el nivel de confianza o su complementario (nivel de significación). Este hecho, y dada la estrecha relación que tienen ambos niveles, los estudiantes pueden confundirlos, o interpretarlos erróneamente. No ocurrió en el estudio.

Otro error también muy significativo es no saber interpretar la confianza (Respuesta de AML).

El alumno tiene menos confusión en la construcción, al darse el intervalo de confianza como una “formulita” el estudiante la aplica y punto. A la hora de interpretarlo puede tener dudas lo que significa “nivel de significación” (Respuesta de ACR).

CEI6. *Considerar fijos los extremos del intervalo.* Son varios los autores que han descrito errores en la interpretación del significado del intervalo de confianza, como suponer que los extremos del intervalo tienen un valor fijo. Sin embargo, la

interpretación correcta es que los extremos son variables y lo que es constante es el valor desconocido del parámetro (Behar, 2001). Algunos estudiantes de la muestra razonaron de acuerdo a este conflicto semiótico.

Un alumno sabe hacer un intervalo de confianza de modo mecánico, se lo explicas paso a paso y lo realiza, se da cuenta que la media muestral pertenece al intervalo y verifica que lo han realizado bien, pero desconoce su interpretación. (Respuesta de CGG).

CEI7. *Aceptar el intervalo sin comprobar que contiene a la media verdadera.* Otro error frecuente es considerar que el valor del parámetro está dentro de cualquier intervalo de confianza, sin tener en cuenta que únicamente un porcentaje (dado por el coeficiente de confianza) de estos lo contiene (Cumming et al., 2004). En la siguiente respuestas se refleja este tipo de error. Se recuerda que algún participante indicó que es posible comprobar si el parámetro pertenece al intervalo, lo cual es un conflicto semiótico en la definición del mismo. Ahora los mismos participantes suponen que se comete un error si no se comprueba este punto.

Aceptar sin comprobar que los intervalos de confianza contienen la media verdadera en el 95% de los casos (Respuesta de MMB).

CEI8. *Considerar el parámetro dentro del intervalo.* Por otro lado, algunos participantes en el estudio indican que es un error suponer que el intervalo contiene al parámetro, es decir, interpretarlo en forma determinista, interpretación que también se dio en nuestros estudiantes.

Por último, desde un punto de vista de interpretación, debido a la falta de importancia que se le da al concepto en el ciclo de secundaria, el alumno puede errar a la hora de interpretar el resultado, en cuanto a considerar que el valor de la media, por ejemplo, estará en el intervalo de forma segura. (Respuesta ACG).

CEI9. *Interpretación incorrecta de factores que afectan al resultado.* Por último, y común a los dos métodos puede darse una interpretación incorrecta de los factores que afectan al resultado. Por ejemplo, no tener en cuenta la importancia de la muestra o los factores de los que depende la amplitud del intervalo.

Si se pide que calculen el nivel de confianza o tamaño de la muestra, nos encontramos con: 1) Tiende a confundir error y amplitud de tal forma que cuando se pide que calcule el nivel de confianza o tamaño de la muestra y le damos el dato de amplitud, lo toma como el error; 2) A la hora de calcular el tamaño de la muestra, si el resultado obtenido es decimal suele no redondear o redondear al valor más próximo y no por exceso como es necesario; 3) A la hora de calcular el nivel de confianza, suele responder con el resultado que recoge la tabla normal para $z_{\alpha/2}$ donde $z_{\alpha/2}$ es tal que $P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$.

Es decir, responde con $(1 - \frac{\alpha}{2})\%$ en lugar de con $(1 - \alpha)\%$ (Respuesta de MVC).

En la Tabla 4.7.7 mostramos los errores citados sobre interpretación. El análisis de los mismos pone de manifiesto la importancia que los futuros profesores otorgan a la interpretación y contextualización de los resultados y a la interpretación del intervalo de confianza. Con menor frecuencia se citan el resto de errores; no saber cuándo rechazar o no rechazar y la importancia de cuidar el lenguaje a la hora de concluir evitando emplear la palabra aceptar. De nuevo se citan muchos de los conflictos de interpretación que aparecieron en los propios estudiantes, por lo cual la corrección posterior en clase de los problemas parece haber ayudado a adquirir un conocimiento sobre las dificultades de los estudiantes. También en un par de casos se citan otros errores, uno de ellos, correctamente, pero no el otro, ya que los tipos de error en un contraste son inherentes a este, no se trata de un error de los estudiantes.

Tabla 4.7.7.

Número de Errores Procedimentales Citados (N.) y Porcentaje (respecto al total de futuros profesores)

Código	Categorías	Contraste		Intervalo		Aparece en la muestra
		N.	%	N.	%	
CEI1	No saber cuándo rechazar o no rechazar la hipótesis nula	16	22,9	--	--	Sí
CEI2	Utilización del término “acepta” la hipótesis nula	8	11,4	--	--	No
CEI3	Cometer un error tipo I o un error tipo II	14	20,0	--	--	No
CEI4	No contextualizar el resultado	31	44,3	--	--	Sí
CEI5	Interpretación del nivel de confianza o nivel de significación	--	--	12	17,1	Sí
CEI6	Considerar fijos los extremos del intervalo	--	--	31	44,3	Sí
CEI7	Aceptar el intervalo sin comprobar que contiene a la media verdadera	--	--	6	8,6	Sí
CEI8	Considerar el parámetro dentro del intervalo	--	--	5	7,1	Sí
CEI9	Interpretación incorrecta de factores que afectan al resultado	1	1,4	7	10,0	Sí

4.7.6. Formas de explicar el intervalo de confianza en el aula

Los participantes en el estudio son futuros profesores. Por ello, no solo deben tener conocimiento común del contenido, sino también un conocimiento didáctico en sus

diferentes componentes, es decir, aquél que es propio del profesor y que le ayuda a proponer situaciones didácticas y organizar la enseñanza. Por ello, junto a las variables descritas anteriormente se han analizado las distintas propuestas metodológicas que utilizarían los participantes a la hora de llevar a cabo la explicación de dicho contenido en un aula con estudiantes. En este apartado se han distinguido dos categorías.

CA1. *Uso de un simulador.* Hoy día hay disponibles numerosos recursos informáticos, tanto en Internet como en programas como SPSS o incluso en la calculadora gráfica que permite visualizar la construcción de intervalos de confianza. De hecho, durante la corrección de las soluciones al intervalo de confianza, se trabajó con los futuros profesores utilizando el simulador de intervalos de confianza disponible en la colección de Applets de Allan Rossman y Beth Chance (véase Figura 4.7.1) (<http://www.rossmanchance.com/applets/ConfSim.html>). Dicha visualización permite construir y representar un número dado de intervalos de confianza, con diferentes poblaciones de partida; entre ellas la distribución normal, de la cual se pueden elegir sus parámetros.

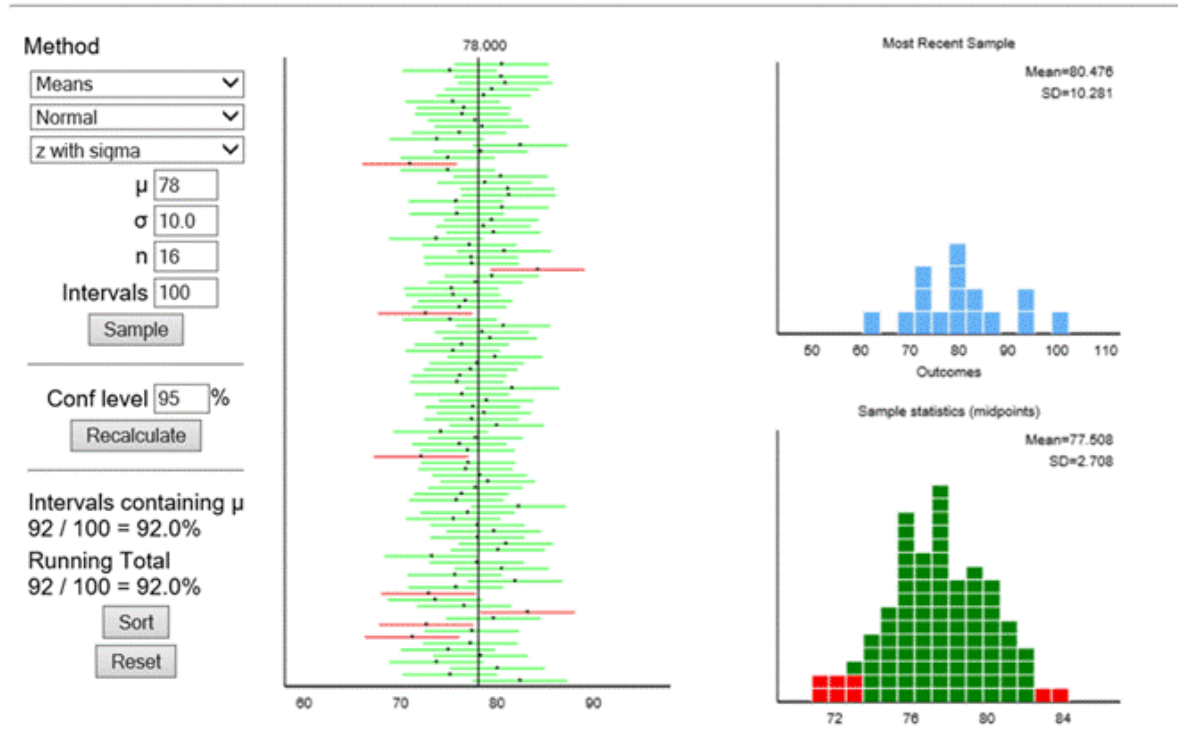
En la Figura 4.7.1 se presenta la simulación llevada a cabo durante la corrección de los ejercicios, tomando como parámetros los propios del problema planteado. El simulador va representando simultáneamente los datos de cada muestra y la distribución muestral del estadístico. El valor verdadero del parámetro se traza sobre una línea vertical y los intervalos que no lo cubren se marcan en color rojo. Por tanto, es posible explorar la variabilidad de los intervalos en diferentes muestras, así como el efecto en el ancho del intervalo de cambios en el tamaño de la muestra o el nivel de confianza, pues todos estos valores pueden cambiarse en el simulador.

En sus respuestas, y recordando la experiencia realizada en la clase, algunos de los participantes en el estudio sugirieron que usarían un simulador para reforzar la comprensión del estudiante y las propiedades asociadas al mismo, como es el caso de la respuesta de PCV.

Para explicarle a un alumno el concepto de intervalo de confianza y sus propiedades, una buena opción sería acudir a un simulador de intervalos de confianza mediante un applet. Introduciendo en ésta los datos correspondientes, el alumno deberá ir tomando sus propias conclusiones, para que de este modo ir evaluando el afianzamiento de los conocimientos que van adquiriendo, y solventando simultáneamente las dudas que vayan surgiendo o se detecten (Respuesta de PCV).

Figura 4.7.1. *Simulador de intervalos de confianza*

Simulating Confidence Intervals



CA2. *Siguiendo un procedimiento mecánico.* Por el contrario, otros participantes únicamente se limitaron a repetir la definición y mostrar el procedimiento de cálculo del intervalo de confianza. Se consideran en esta categoría las respuestas que proponen únicamente la sustitución de los datos facilitados en el problema en la expresión correspondiente del intervalo de confianza, sin ahondar en porqué se lleva a cabo ese procedimiento, ejemplo de ello es la respuesta dada por MGG. Este tipo de enseñanza, a lo que parece frecuente en las aulas no refuerza la comprensión, sino que los estudiantes aprenden reglas rutinarias que no comprenden; por tanto, es fácil que cometan errores.

Pasos a seguir:

1° Definir datos;

2° Calculamos $z_{\alpha/2}$;

3° Aplicamos la fórmula del intervalo de confianza (Respuesta de MGG).

El análisis frecuencial de las respuestas en relación a las distintas propuestas metodológicas nos ha permitido comprobar las dificultades que tienen los futuros profesores para explicar a sus futuros estudiantes el intervalo de confianza. Solamente un 40% han propuesto alguna estrategia que incluya la simulación para reforzar el conocimiento conceptual o calculadoras con el fin de que se obtenga fácilmente los extremos del intervalo por parte de sus estudiantes. Aun así, destacamos que la mayoría de las aportaciones que se hacen (25 de 29) están enfocadas en aplicar de una forma mecánica el procedimiento sin distinguir los elementos principales y más destacados de un intervalo de confianza. Por otro lado, el resto de los participantes ni siquiera indicaron una posible forma de enseñar el intervalo de confianza.

Los participantes del estudio están preparándose para ser futuros profesores en centros de Enseñanza Secundaria Obligatoria y Bachillerato, y dicho contenido forma parte tanto en la asignatura de Matemáticas Aplicadas a las Ciencias Sociales II de segundo curso de bachillerato (MECD, 2015) como de las pruebas de acceso a la universidad. Estos resultados son altamente preocupantes ya que los participantes muestran un conocimiento muy pobre de la faceta mediacional de su conocimiento sobre los intervalos de confianza.

4.7.7. Síntesis de conocimientos didácticos

La representación gráfica del número de errores que cita cada participante sobre contraste de hipótesis indica que lo más frecuente es citar cinco errores posibles de sus futuros estudiantes (véase Figura 4.7.1), aunque alguno cita hasta diez. Asimismo, el gráfico correspondiente a la frecuencia acumulada refleja que son pocos los que citan más de seis.

En el caso del intervalo de confianza se observa que es más frecuente citar 3 errores por cada participante (véase Figura 4.7.2), aunque alguno cita hasta siete. Asimismo, la gráfica de la frecuencia acumulada revela que son pocos los que citan más de cinco. En definitiva, si comparamos los resultados de ambas gráficas se observa que en general los estudiantes recuerdan más propiedades del contraste de hipótesis que del intervalo, posiblemente porque este procedimiento implica más conceptos diferentes que el intervalo.

Figura 4.7.1. Distribución y distribución acumulada (en porcentaje) del número de errores citados por participante sobre contraste de hipótesis

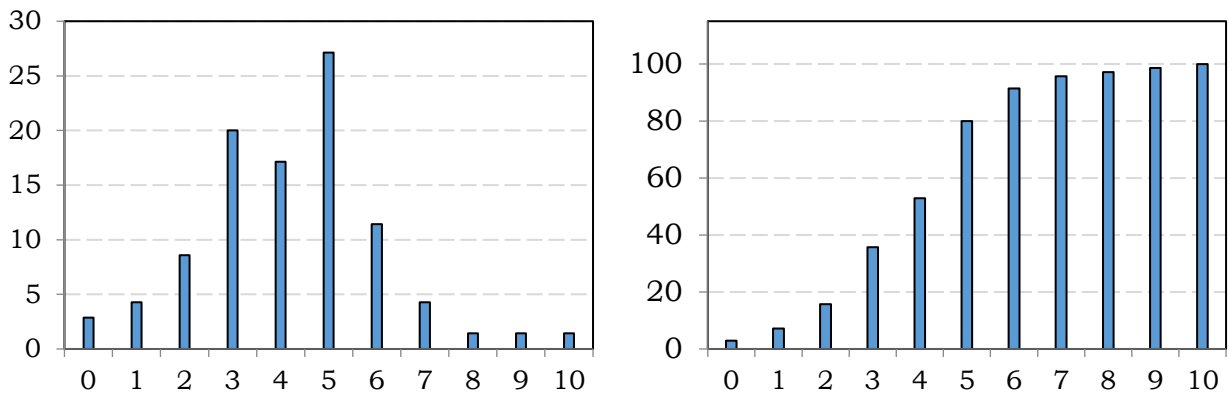
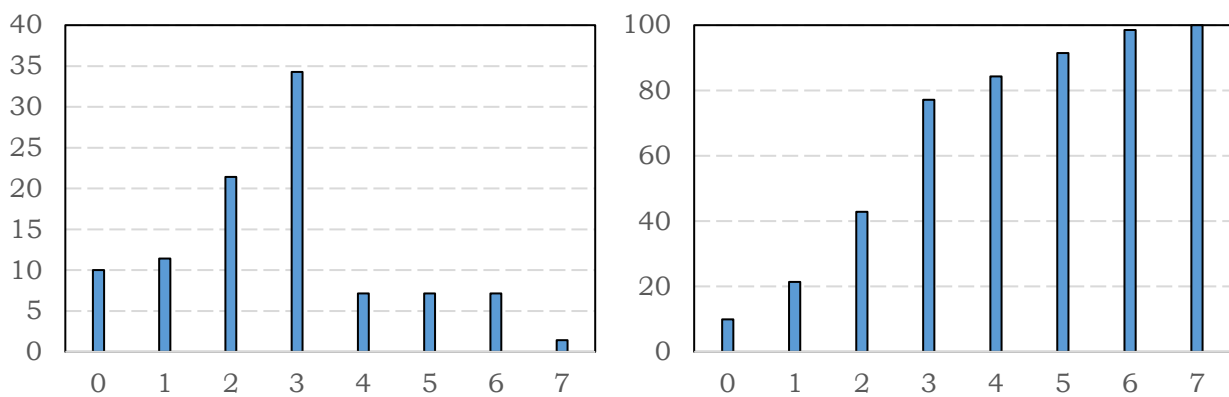
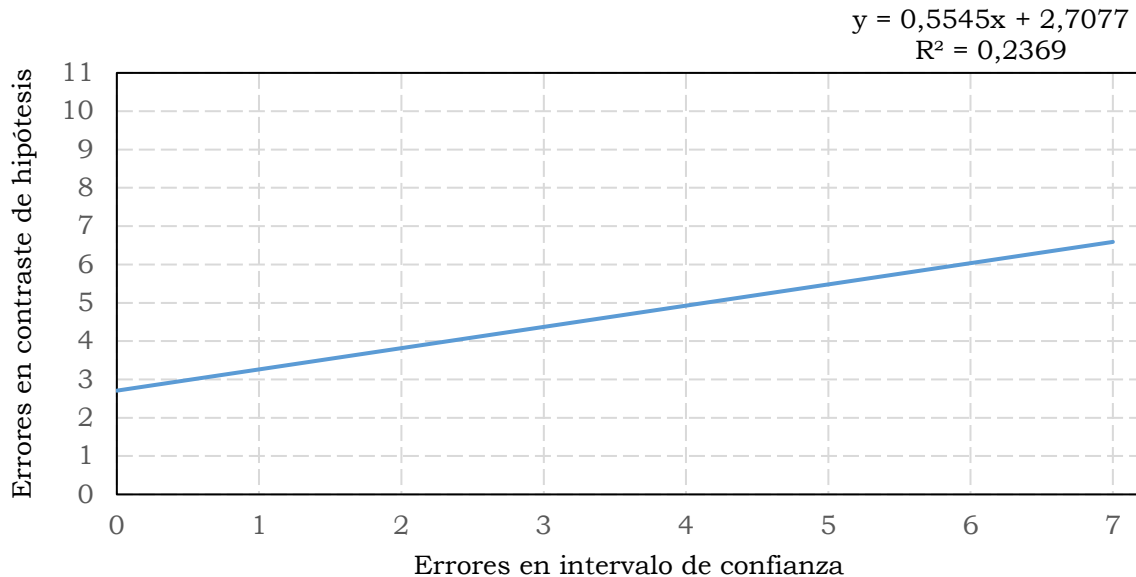


Figura 4.7.2. Distribución y distribución acumulada (en porcentaje) del número de errores citados por participante sobre intervalo



Una cuestión a estudiar es si existe relación entre el número de errores citados por el mismo participante respecto a ambos procedimientos. Para comprobarlo, en la Figura 4.7.3 se representa el diagrama de dispersión del número de errores citados en el contraste en función del número citado en relación al intervalo de confianza. Se observa una correlación positiva y moderada ($r = 0,49$; $R^2 = 0,2369$), lo que sugiere que hay una tendencia moderada a que el estudiante que recuerda más errores de uno de los dos procedimientos también lo recuerda del otro. Ello indica que, conjuntamente las dos preguntas nos han permitido evaluar aspectos del conocimiento de la faceta cognitiva de la inferencia, pues el número total de errores en los dos apartados daría una medida de este conocimiento.

Figura 4.7.3. *Relación entre número de errores citados para el intervalo de confianza y el contraste de hipótesis*



4.7.8. Conclusiones sobre el conocimiento didáctico sobre inferencia estadística

Para finalizar el estudio del conocimiento didáctico de los futuros profesores, incluimos en este apartado nuestras conclusiones sobre sus resultados.

Los participantes del estudio fueron capaces de identificar un número apreciable de errores en el desarrollo del contraste de hipótesis e intervalo de confianza por parte de sus futuros estudiantes. La mayoría de los errores señalados coinciden con los identificados en la literatura; por ejemplo, confundir hipótesis nula y alternativa, confundir estadístico o parámetro o plantear hipótesis no complementarias (Batanero, 2001; Vallecillos, 1999). Igualmente se alude a la confusión entre contraste unilateral o bilateral, confusión de la distribución muestral e interpretación de resultados, encontrados en Espinel et al. (2007). Asimismo, se identifican errores de tipificación y lectura de la tabla normal que no son tan específicos del contraste de hipótesis pero que ocurren con frecuencia entre los estudiantes en este y otros temas.

Sin embargo, muchos otros errores, incluso cometidos por los propios participantes y corregidos en la clase, se olvidan, en particular, diferentes interpretaciones incorrectas del p -valor. En muchos casos se describen los errores con poca precisión, lo que nos indica que en realidad su conocimiento del tema es escaso;

por ejemplo, cuando se alude a hipótesis mal planteadas, interpretación incorrecta del intervalo de confianza o no saber plantear bien el contraste. También se describen errores poco frecuentes como la confusión en el tipo de error, o se describen como errores de los estudiantes los errores asociados al procedimiento (error tipo I y II) que son inevitables.

En consecuencia, aunque se observa un cierto grado de desarrollo de la faceta cognitiva del conocimiento didáctico-matemático de estos profesores (en el modelo de Godino, 2009) en lo que se refiere a la inferencia estadística, sería necesario desarrollar de una forma más completa dicha faceta en estos futuros profesores haciéndoles conocer mejor el desarrollo de los procedimientos de contraste de hipótesis e intervalo de confianza y los posibles errores de los estudiantes. Además, sería necesario desarrollar también en ellos el resto de las facetas de su conocimiento didáctico-matemático sobre este tema.

Los resultados obtenidos en el análisis de cómo explicarían el intervalo de confianza en el aula pone en relieve la necesidad de desarrollar su faceta instruccional para proporcionarle modelos y situaciones didácticas que ayuden a mejorar el conocimiento de los estudiantes sobre el contraste de hipótesis y evitar en el futuro los errores de aplicación descritos en los antecedentes. En este sentido sería también deseable formarles en métodos alternativos de aproximarse al contraste de hipótesis. Como se indica en Batanero et al. (2017) no hay una única aproximación al contraste de hipótesis (en general a la inferencia estadística) y es posible hoy día, gracias a la simulación, comenzar con aproximaciones informales. Todo ello teniendo en cuenta que una aproximación informal por sí sola no asegura que los estudiantes no cometan errores conceptuales.

4.8. CONCLUSIONES DEL ESTUDIO 2

En base a los objetivos e hipótesis establecidas en la Sección 4.2 y Sección 4.3 respectivamente, se desarrollan las conclusiones más destacadas obtenidas en el Estudio 2.

4.8.1. Conclusiones sobre los objetivos

Los objetivos de este estudio desarrollan el tercer objetivo general de la investigación, que era el siguiente y que se justificó por la ausencia de trabajos sobre este tema:

OE3. Evaluar algunos elementos del conocimiento didáctico-matemático sobre la inferencia en futuros profesores de Educación Secundaria Obligatoria y Bachillerato.

Dicho objetivo se desglosó en tres objetivos parciales que son analizados a continuación:

OE3.1. Evaluar el conocimiento matemático común y avanzado sobre el contraste de hipótesis en futuros profesores de Educación Secundaria Obligatoria y Bachillerato.

Para conseguir este objetivo se planteó un problema abierto, similar a los propuestos en años anteriores en las PAU, como parte de un taller formativo dirigido a desarrollar el conocimiento didáctico-matemático (Godino et al., 2017; Pino-Fan y Godino, 2015) sobre inferencia estadística que se llevó a cabo en el marco del Máster de Formación del Profesorado de Educación Secundaria Obligatoria y Bachillerato, especialidad de Matemáticas

El análisis de los resultados mostrados en la Sección 4.5 nos ha permitido ver que, aunque los futuros profesores en general establecen las hipótesis del contraste de hipótesis adecuadamente y eligen un procedimiento pertinente para las hipótesis planteadas, pocos de ellos desarrollan correctamente todos los pasos del contraste e interpretan y contextualiza los resultados obtenidos. A partir de los errores cometidos, similares a los encontrados en la investigación previa (Birnbaum, 1982, Cañadas et al., 2012, Espinel et al., 2007; Krauss y Wassner, 2002; Saldanha y Thompson, 2007; Vallecillos, 1994) se identificaron una serie de conflictos semióticos conceptuales, procedimentales y representacionales sobre este contenido.

Se ha notado poca precisión cuando se describen algunos de los términos empleados en la elaboración del contraste. Un ejemplo de ello sería el análisis realizado en el desarrollo del procedimiento donde el p -valor es obtenido en algunos casos como resultado de una probabilidad simple o como una probabilidad puntual. Igualmente,

destacamos el hecho del escaso número de futuros participantes que llevan a cabo la interpretación y contextualización de los resultados obtenidos. Recordemos que las normativas curriculares señalan la importancia de que los estudiantes deben realizar interpretaciones de estudios estadísticos.

En resumen, el estudio sobre el conocimiento matemático en relación al contraste de hipótesis pone de manifiesto las dificultades que presentan los futuros profesores cuando abordan un problema sobre contraste de hipótesis. Se ha observado que un alto porcentaje de los participantes no han realizado correctamente el ítem analizado e incluso, seis de estos no han sido capaces de llevar a cabo ninguna de las etapas en las que dividido el desarrollo del contraste.

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

OE3.2. Evaluar el conocimiento matemático común sobre el intervalo de confianza en futuros profesores de Educación Secundaria y Bachillerato.

Este objetivo se ha llevado a cabo mediante el análisis de las respuestas de los futuros profesores al ítem 2, similar al planteado a los estudiantes de Bachillerato en las PAU presentados en la Sección 4.6. Los resultados obtenidos muestran que, aunque la mayoría de los participantes en el estudio plantea y calcula correctamente los extremos del intervalo de confianza, menos de la mitad interpretan y definen correctamente el término en cuestión.

Asimismo, las conclusiones obtenidas en el punto de síntesis del conocimiento nos han permitido ver la dificultad que entraña una correcta elaboración del intervalo de confianza por parte de los participantes ya que más del 70% han tenido dificultades

durante el proceso de ejecución. De nuevo en este apartado se han identificado y clasificado una serie de conflictos semióticos, algunos de los cuáles se han descrito como errores en la investigación previa (por ejemplo, en Behar, 2001; Cumming y Fidler, 2005; Fidler y Cumming, 2005 y Olivo, 2007).

OE3.3. Evaluar algunos elementos de la faceta cognitiva del conocimiento didáctico-matemático de los futuros profesores de Educación Secundaria Obligatoria y Bachillerato. En concreto, se desea evaluar el conocimiento que los futuros profesores de Educación Secundaria Obligatoria y Bachillerato tienen de los posibles errores de los estudiantes en el contraste de hipótesis e intervalo de confianza.

De la información recogida a lo largo del capítulo y resumida en la Tabla 4.7.1 se observa que los futuros profesores una variedad de errores que podrían cometer los estudiantes cuando realizan un problema de contraste de hipótesis o intervalo de confianza. Estos resultados muestran un cierto conocimiento de los errores asociados a este tipo de problemas. Destacamos la importancia que se le asigna a los errores de tipo procedimental (30,4% en el contraste de hipótesis y 48,7% en el intervalo de confianza) y a la interpretación de resultados (23,6% y 31,3% respectivamente). Hay poca consciencia de los errores en el planteamiento de las hipótesis o del procedimiento. Por otro lado, entre algunas de las respuestas analizadas se han encontrado muchas imprecisas las cuales podría esconder falta de conocimiento sobre las mismas.

Aun así, es necesario que los futuros profesores profundicen más en los distintos errores que pueden cometer los estudiantes ya que el desarrollo de una adecuada faceta cognitiva ayuda a afrontar de una forma más adecuada el desarrollo del tema en el aula.

Los resultados fueron peores cuando se le ha preguntado por las propiedades fundamentales que recordaban, pues menos de un cuarto han nombrado alguna propiedad en relación a los intervalos de confianza y se han citado más errores procedimentales que conceptuales, lo que indica que los futuros profesores no son conscientes de la dificultad conceptual del tema.

Por otro lado, el estudio de las propuestas metodológicas que realizan los futuros profesores con el fin de abordar el tema dentro del aula revela las limitaciones que estos tienen, pues únicamente un tercio de ellos señalan alguna. En general, se decantan por

una enseñanza tradicional y procedimental, sin tener en cuenta los recursos de cálculo y de simulación, incluso ni siquiera los que ellos mismos utilizaron a lo largo de las sesiones formativas.

En resumen, el conocimiento evaluado sobre algunos aspectos de las componente cognitiva y mediacional sobre la inferencia es muy escaso en la muestra de estudiante y estos resultados podrían ser similares en otros grupos de futuros profesores. Ello implica la necesidad de reforzar la preparación didáctica de los futuros profesores para lograr una enseñanza razonable de la inferencia estadística en Bachillerato. Junto a un conocimiento matemático del tema en cuestión, es necesario formar a los futuros formadores en temas tan importantes como la didáctica con el fin de garantizar el éxito de aprendizaje en los estudiantes.

4.8.2. Conclusiones sobre las hipótesis

De los objetivos señalados, así como del estudio de los antecedentes, se deducen algunas hipótesis que podríamos analizar a partir de los resultados de la evaluación y que son las siguientes:

H3.1. Al evaluar los conocimientos de los futuros profesores se prevé encontrar algunas dificultades similares a las descritas en estudiantes en conceptos estadísticos requeridos para el trabajo con la inferencia que deben enseñar, esto es, en su conocimiento común del contenido.

El análisis de los diferentes conflictos semióticos identificados en las respuestas de los estudiantes aporta evidencias a favor de esta hipótesis. Muchos de ellos han sido recogidas como errores en la literatura.

Congruentemente con la literatura previa, tanto la determinación del p -valor en un contraste de hipótesis (Batanero, 2000; Birnbaum 1982; Falk, 1986; Espinel et al., 2007) como la interpretación del nivel de confianza (Behar, 2001; Cumming et al., 2012; Olivo, 2008) suponen todavía un reto para los participantes del presente estudio. En menor medida, se ha detectado errores en identificar el papel que juega la hipótesis nula y la hipótesis alternativa en el planteamiento del contraste (Batanero et al., 2012; Cañadas et al., 2012; Vallecillos, 1999), la discriminación entre estadística y parámetro (Harradine et al., 2011; Schuyten, 2001; Olivo, 2008) y en el planteamiento del intervalo

de confianza.

En general, se ha percibido que gran parte de las respuestas dadas por los futuros profesores estaban centradas en la realización de las tareas mediante un procedimiento mecanizado y sin contextualizar el significado a cada uno de los elementos que intervienen en el desarrollo. Este hecho adquiere cierta importancia ya que implica un desconocimiento de la esencia que existe en la realización de un contraste o determinación de los extremos de un intervalo.

H3.2. Se espera también encontrar algunas dificultades al emplear procedimientos estadísticos básicos, requeridos en la aplicación de la inferencia, es decir en la parte procedimental de su conocimiento común de la inferencia estadística.

De igual manera, los resultados obtenidos en el presente estudio muestran que futuros profesores con una formación académica en matemáticas e ingeniería presentan un alto porcentaje de errores en algunos de los procedimientos básicos. Estos resultados son congruentes con trabajos previos sobre el uso de las tablas estadísticas y la tipificación de variables aleatorias continuas (Alvarado, 2007, Batanero y Vallecillos, 1994). Los participantes recurren al uso de la propiedad de la distribución de probabilidad en un entorno de la media o aproximan a cero en el cálculo del p -valor. Es decir, muestran dificultades en la tipificación y el uso de las tablas de la distribución normal $N(0,1)$. Por otro lado, se han encontrado errores en la obtención de los cuantiles que determinan las regiones de rechazo en un contraste de hipótesis o que son empleados para la obtención de los intervalos de confianza. Pensamos que ello es debido a la confusión en la aplicación de desigualdades o la determinación de las regiones de un contraste (Chance et al., 2004; Espinel et al., 2007; Inzunza, 2006).

Asimismo, señalamos que, aunque algunos expresan de forma detallada el proceso de elaboración de un contraste de hipótesis o de la determinación de un intervalo de confianza, termina la elaboración del ítem sin realizar los cálculos necesarios con objeto de tener la información solicitada en el mismo. O incluso cometen errores de cálculo.

H3.3. Los futuros profesores tendrán más dificultad en la interpretación de los resultados de la inferencia que en el planteamiento de los procedimientos o en la realización de los cálculos asociados.

La interpretación y contextualización de los resultados obtenidos después de llevar a cabo un contraste de hipótesis o una vez determinados los extremos de un intervalo de confianza supone un impedimento ya que sin una interpretación y/o contextualización adecuada se pierde la esencia de porqué se elabora este tipo de problemas. Aunque los participantes señalan que un estudiante puede cometer errores de interpretación (véase Sección 4.7), los resultados recogidos a lo largo de la Sección 4.5 y Sección 4.6, dejan en evidencia que incluso los futuros profesores comenten errores cuando tratan de concluir con los resultados obtenidos. Por ejemplo, aceptan una hipótesis cuando deben rechazar o no llegan a la decisión o bien deciden rechazar la hipótesis, pero no contextualizan.

En consecuencia, y al igual que ocurrió en el estudio de Arteaga (2011), a los participantes les resulta difícil la interpretación de los resultados de los procedimientos inferenciales. Por ello no completan el ciclo de modelización de Chaput et al. (2011): abstraen la realidad (esperanza de vida y los datos del problema) para plantear las hipótesis y de ahí pasar a un modelo matemático adecuado, con el que trabajan, pero no logran el paso final, consistente en traducir dicha conclusión a la realidad estudiada (esperanza de vida media al nacer). Tampoco ejercitan el proceso de integración de estadística y contexto descrito por Wild y Pfannkuch (1999).

H3.4. Los futuros profesores de Educación Secundaria Obligatoria y Bachillerato desconocen parte de los errores de aplicación e interpretación de la inferencia estadística que se han descrito en la investigación previa, mostrando carencias en la faceta cognitiva de su conocimiento didáctico-matemático.

Pese a que los participantes han listado un gran número de errores procedimentales y señalan la importancia de realizar una interpretación correcta de los resultados, destacamos que muchos de ellos se han definido de una forma muy imprecisa indicándolos de forma muy superficial. Por otro lado, no citaron algunos de los errores asociados a la interpretación del nivel de significación y del p -valor, ampliamente citados en la investigación previa, por ejemplo, en Batanero (2000) y Caperos y Pardo (2013). Además, ellos mismos cometieron algunos de estos errores al resolver los problemas; por ejemplo, intercambiaron los dos términos de la probabilidad condicional que define el p -valor. Todo ello nos hace sospechar que los participantes pudieran no ser conscientes de los errores relacionados con la interpretación de estos

conceptos.

El número de errores procedimentales y de interpretación de resultados sobrepasa bastante a los conceptuales o de planteamiento. Por un lado, esto indica que los futuros profesores consideran estos puntos más difíciles, incluso aunque hoy día los errores procedimentales desaparecen prácticamente con el uso del software estadístico. Pero, por otro lado, este resultado sugiere que los futuros profesores dan mayor peso al aprendizaje de los procedimientos de cálculo que a la comprensión profunda de los objetos matemáticos involucrados en los mismos. Ello a pesar de su complejidad conceptual, especialmente del contraste de hipótesis, señalada en Inzunza y Jiménez (2013) y de sus propios errores conceptuales en el desarrollo del contraste de hipótesis,

La mayoría de los errores señalados por los futuros profesores aparecieron en su propia solución de los problemas; de lo que se deduce que la actividad posterior de discusión de las soluciones en grupo fue útil para desarrollar el conocimiento de los futuros profesores sobre los errores que pueden cometer sus estudiantes.

Además, prácticamente todos los errores que describen coinciden con los identificados en la literatura; por ejemplo, confundir hipótesis nula y alternativa, confundir estadístico y parámetro o plantear hipótesis no complementarias (Batanero, 2000; Vallecillos, 1999). Igualmente se alude a la confusión entre contraste unilateral o bilateral, confusión de la distribución muestral e interpretación de resultados, encontrados en Espinel et al. (2007). En la misma línea de las investigaciones de Fidler y Cumming (2005), los participantes destacan los errores asociados a la interpretación del intervalo de confianza. Asimismo, en ambas tareas se identifican errores de tipificación y lectura de la tabla normal, descritos por Ramos et al. (2009) y la confusión entre estadístico y parámetro (Harradine et al., 2011).

Estos resultados ponen de manifiesto la necesidad de formar a los futuros profesores en este tema ya que un desarrollo adecuado de la faceta cognitiva es positivo pues un alto conocimiento de esta información facilita la mejora del proceso de enseñanza del tema en cuestión. Todo ello sin olvidar la atención a las actitudes del profesorado (Estrada et al., 2011) que con frecuencia se sienten poco preparados para la enseñanza del tema.

Capítulo 5

CONCLUSIONES GENERALES

Índice de Contenidos

- 5.1. Introducción
 - 5.2. Conclusiones respecto a los objetivos del trabajo
 - 5.3. Conclusiones sobre las hipótesis
 - 5.4. Principales aportaciones
 - 5.5. Limitaciones del estudio y futuras líneas de investigación
 - 5.6. Implicaciones para la enseñanza
-

5.1. INTRODUCCIÓN

Para finalizar esta memoria, se dedica este capítulo a la exposición de las principales conclusiones obtenidas en la investigación.

En primer lugar, se analizan cada uno de los objetivos generales del trabajo, expuestos en el Capítulo 1, para deducir hasta qué punto se ha logrado alcanzar cada uno de ellos, completando así el análisis de los objetivos específicos de los diferentes estudios realizados, estudiados en cada uno de los capítulos anteriores. Igualmente, se analizan las hipótesis generales planteadas en el Capítulo 1, que se desglosaron en hipótesis parciales en cada uno de los capítulos anteriores.

Se destacan, a continuación, las principales aportaciones de esta investigación, así como sus limitaciones, cuyo análisis nos permite plantear posibles líneas de trabajo para continuar el expuesto en la memoria.

Se finaliza el capítulo con algunas sugerencias para la enseñanza de la inferencia y la formación de los profesores que se deducen de nuestra investigación.

5.2. CONCLUSIONES RESPECTO A LOS OBJETIVOS DEL TRABAJO

Tal y como ha sido señalado en la introducción de esta memoria, la investigación desarrollada en la misma se orientó, en primer lugar, a caracterizar la presencia de la inferencia estadística en el currículo de Bachillerato y en los problemas propuestos en las pruebas de acceso a la Universidad, determinando las principales variables que los caracterizan. Con ello se continuaba nuestro trabajo previo analizando los problemas de probabilidad en dichas pruebas (Batanero, López-Martín, Arteaga y Gea, 2018; Carretero et al., 2015; Contreras et al., 2015).

Una segunda parte de la investigación estuvo centrada en evaluar algunos elementos del conocimiento didáctico-matemático (Godino, 2009; Godino et al., 2017; Pino-Fan y Godino, 2015) de los futuros profesores de matemáticas de Educación Secundaria Obligatoria y Bachillerato en relación con la inferencia.

De estos dos objetivos principales del trabajo se desglosaron los siguientes objetivos específicos:

OE1. Realizar una síntesis de la investigación sobre dificultades en la comprensión y aplicación de la inferencia estadística y del análisis de los problemas propuestos en las pruebas de acceso a la universidad. Previa a los estudios empíricos propuestos, se consideró analizar las investigaciones previas, para fijar nuestro trabajo en un contexto más amplio.

La realización de esta síntesis fue necesaria para completar las realizadas por otros autores, como Batanero y Díaz-Batanero (2006), Castro-Sotos et al. (2007), Harradine et al. (2011), Tobias-Lara y Gomez-Blancarte (2019).

Pensamos que se ha logrado razonablemente este objetivo, mediante la síntesis bibliográfica expuesta en el Capítulo 2. En dicho capítulo se expone el resumen del

trabajo de síntesis en que se localiza, clasifica y analiza un gran número de investigaciones en relación con diferentes objetos matemáticos que forman parte de la inferencia estadística. Además, se ha realizado una interpretación de dichos errores en términos de conflictos semióticos (Godino, 2002; 2019; Godino y Batanero, 1994). Los resultados se han utilizado para comparar con las respuestas de los futuros profesores en el Estudio 2 e interpretar dichas respuestas en términos de conflictos semióticos.

Del mismo modo se ha incluido otra síntesis de investigaciones sobre los contenidos de inferencia estadística en las pruebas de acceso a la universidad y en libros de texto. Esta parte de la revisión ha permitido describir las variables a analizar en el Estudio 1.

El interés de este objetivo radica también en que el conocimiento sobre los errores más frecuentes o dificultades que tienen los estudiantes cuando trabajan los contenidos de inferencia estadística, supone una ayuda a los docentes. El hecho de contar con esta información permite a los profesores adaptar su proceso de enseñanza-aprendizaje con el fin de que los estudiantes asimilen, de una forma correcta, los contenidos incluidos en el bloque objeto de estudio.

OE2. Caracterizar los problemas de inferencia propuestos en las pruebas de acceso a la Universidad y comparar el significado evaluado en dichas pruebas con el significado institucional pretendido para la inferencia en las orientaciones curriculares.

Este objetivo se justificó porque los problemas propuestos a los estudiantes, bien durante la enseñanza, o en las pruebas de evaluación, van a incidir directamente en el aprendizaje de diferentes objetos matemáticos y el significado asignado a los mismos. Además, por la importancia que se da en nuestro marco teórico a las situaciones-problema, pues las prácticas involucradas en su resolución son el origen de los objetos matemáticos (Godino, 2002; 2017; Godino et al., 2007, 2020).

Para lograr este objetivo y en base al enfoque Onto-Semiótico, se han resuelto cada uno de los problemas las pruebas PAU para Matemáticas Aplicadas a las Ciencias Sociales II durante el periodo 2003-2014 con el fin de analizar los objetos relacionados con la inferencia estadística y determinar el significado evaluado en las pruebas. Igualmente se han analizado diferentes variables que inciden en la dificultad de los

problemas, analizando la distribución de sus modalidades a lo largo del tiempo. Esto nos ha permitido en el Capítulo 3 presentar las prácticas matemáticas y configuraciones epistémicas ligadas a diferentes campos de problema y analizar el significado institucional de la inferencia evaluado en las pruebas de acceso a la universidad.

Este significado evaluado se ha comparado con el significado institucional pretendido en las orientaciones curriculares para Bachillerato y el estudio se complementa con el análisis de algunos aspectos de la idoneidad didáctica de la inferencia en las PAU. La información obtenida de este estudio será de utilizad para los profesores que enseñan la inferencia estadística en Bachillerato y preparan a sus estudiantes para enfrentarse a las pruebas de acceso a la universidad.

OE3. Evaluar algunos elementos del conocimiento didáctico-matemático sobre la inferencia estadística en futuros profesores de Educación Secundaria Obligatoria y Bachillerato. Más concretamente, se desea evaluar algunos elementos de su conocimiento común del contenido y de la faceta cognitiva de su conocimiento didáctico (Godino et al., 2017; Pino-Fan y Godino, 2015).

En el estudio de los antecedentes, presentado en el Capítulo 2, se justificó la escasez de trabajos de evaluación de los conocimientos matemáticos sobre inferencia estadística de los futuros profesores de Educación Secundaria Obligatoria y de Bachillerato que se han reducido a evaluaciones puntuales de sus conocimientos, sin profundizar en el análisis de las soluciones de los futuros profesores a problemas elementales de inferencia estadística. Respecto al conocimiento didáctico no se encontraron antecedentes en el tema.

En consecuencia, para cumplir este objetivo, basándose en los estudios anteriores, se preparó un cuestionario de evaluación de algunos elementos del conocimiento de los futuros profesores, cuya necesidad se justifica por no haber encontrado uno en los antecedentes que sirviese para nuestros objetivos.

En el Capítulo 4 se analiza el contenido de dicho cuestionario y los resultados de la evaluación, que se llevó a cabo, dentro de un taller de inferencia a 73 futuros profesores en el Máster de Educación Secundaria y Bachillerato, especialidad de Matemáticas. Al haber elegido profundizar más en las respuestas que realizar una

evaluación de un contenido más amplio, nos limitamos a proponer dos problemas y algunas preguntas sobre la didáctica del tema en el cuestionario.

El análisis de las respuestas de los participantes permitió estudiar, por un lado, sus conocimientos sobre el contraste de hipótesis y el intervalo de confianza, desglosado en diferentes componentes de la solución a los dos problemas planteados. Estas componentes incluyen el planteamiento de hipótesis, selección de un intervalo o contraste apropiado, cálculos realizados para llevarlo a cabo e interpretación de los resultados. Aunque las respuestas fueron mejores que las observadas en investigaciones previas con estudiantes, se encontraron numerosos errores, bien citados en los antecedentes de nuestro trabajo o nuevos. Estos errores han sido analizados e interpretados como conflictos semióticos cognitivos conceptuales o procedimentales en el Capítulo 4. Por otro lado, aunque con menor profundidad, también se ha llevado a cabo la evaluación de algunos elementos del conocimiento didáctico de estos participantes. En concreto, su conocimiento de los errores y dificultades de los estudiantes (faceta cognitiva) y la forma en que enseñarían el intervalo de confianza (faceta mediacional).

Los resultados obtenidos en los dos puntos anteriores se describen con detalle en el Capítulo 4 y constituyen una aportación original del trabajo, que han dado origen a varias publicaciones (Batanero y López-Martín, en prensa; Batanero et al., 2019 y López-Martín et al., 2019) y diferentes presentaciones en congresos. Dichos resultados sugieren la necesidad de reforzar la formación matemática y didáctica de los profesores de matemáticas, para que la enseñanza de la inferencia en Bachillerato se lleve a cabo con éxito.

5.3. CONCLUSIONES SOBRE LAS HIPÓTESIS

En segundo lugar, se analizan las hipótesis generales que se expusieron en el Capítulo 1, entendidas como expectativas sobre lo que se esperaba encontrar en la investigación.

H1. El análisis de las investigaciones previas mostrará que el tema que nos ocupa ha sido escasamente investigado y, por tanto, nuestro trabajo puede aportar información original.

Esta hipótesis se ha visto confirmada en el estudio de antecedentes expuesto en el Capítulo 2. Respecto al análisis de los problemas de inferencia estadística propuestos en las pruebas de acceso a la universidad, encontramos muy escasas investigaciones previas y las halladas, no realizaron un estudio sistemático de las variables que caracterizan dichos problemas. Tampoco llevaron a cabo un análisis semiótico de los campos de problemas que pueden identificarse.

Respecto al estudio de evaluación, las pocas investigaciones previas realizadas con profesores (Behar 2001; Krauss y Wassner, 2002; Yáñez y Behar, 2009) se han centrado en el muestreo o bien abordan únicamente algunos aspectos del contraste de hipótesis. Pero ninguna analiza la forma en que los profesores resuelven problemas similares a los planteados a los estudiantes en las pruebas de acceso a la universidad. Tampoco describen los resultados de la evaluación en términos de conflictos semióticos. Finalmente, no hemos encontrado ningún trabajo previo sobre el conocimiento didáctica de futuros profesores en relación a la inferencia.

H2. El análisis semiótico de los problemas de inferencia propuestos en las PAU y su comparación con el contenido curricular de estadística en los decretos curriculares de Bachillerato mostrará, por un lado, que se da una alta importancia a la inferencia en las PAU. Además, se observará un buen ajuste entre los contenidos de inferencia fijados en el currículo y los considerados en dichas pruebas.

Para estudiar esta hipótesis, en el Capítulo 3 se presenta el estudio desarrollado sobre los problemas de inferencia propuestos en las pruebas de acceso a la universidad. Una de las conclusiones obtenidas fue la alta importancia dada a la inferencia estadística en las pruebas, ya que, uno de los cuatro problemas propuestos para las Matemáticas Aplicadas a las Ciencias Sociales es de inferencia. Esta proporción es bastante mayor que el papel dado a la inferencia estadística en el currículo, pues es solo una parte del bloque sobre Estadística y Probabilidad, que es uno de los cuatro que figura en el currículo. Teniendo en cuenta que las pruebas incluyen también un problema de probabilidad condicional o teorema de Bayes, la mitad de la prueba de evaluación se dedica a Estadística y Probabilidad, que tiene un peso doble en la pruebas al que se señala en el currículo.

Por otro lado, el análisis semiótico de un ejemplo de cada campo de problema y

la identificación de las configuraciones epistémicas asociadas permitió comparar el significado institucional pretendido de la inferencia en las orientaciones curriculares y el evaluado en las pruebas. Se mostró en este análisis un buen ajuste entre los dos tipos de significados.

H3. Los estudios de evaluación llevados a cabo con futuros profesores mostrarán en general una buena preparación matemática sobre el tema. Sin embargo, esperamos encontrar errores conceptuales algunos conflictos semióticos cognitivos relacionados con el tema y sobre todo de interpretación de los conceptos implicados en la inferencia estadística. Además, esperamos mostrar un conocimiento didáctico insuficiente de dichos profesores en las facetas evaluadas en nuestro estudio.

Parte estudiar esta hipótesis se llevó a cabo el estudio de evaluación recogido en el Capítulo 4. La hipótesis se confirma solo en parte. Respecto al contenido matemático, aunque la mayoría de los futuros profesores de la muestra supo plantear adecuadamente y calcular el contraste de hipótesis y el intervalo de confianza pedido, fueron pocos los que les dieron una interpretación adecuada.

Además, en algunos estudiantes se encontraron conflictos semióticos cognitivos conceptuales y procedimentales. Dichos conflictos se relacionan con el planteamiento de hipótesis, diferencia entre estadístico y parámetro, distribución muestral, p -valor y proceso de contraste. Fueron menos los conflictos relacionados con el intervalo de confianza, pero se recuerdan pocas propiedades de mismo.

Estas carencias de conocimiento matemáticos se reflejan también en las componentes evaluadas de su conocimiento didáctico. Los participantes en el estudio fueron capaces de identificar pocas dificultades de sus estudiantes y sugirieron en general, una enseñanza formal y procedimental del tema. Será entonces necesario reforzar la formación didáctico-matemática de los futuros profesores sobre inferencia.

5.4. PRINCIPALES APORTACIONES

Como se ha argumentado a lo largo de la memoria, los puntos investigados en este trabajo han sido escasamente trabajados en la investigación previa. Por tanto, la presente investigación aporta diferentes resultados al área de didáctica de la matemática:

- Por un lado, el análisis semiótico de los campos de problemas de inferencia y la identificación de las principales variables que caracterizan los problemas de inferencia y sus categorías puede servir de base para el análisis de otros problemas de inferencia en pruebas de evaluación o en los textos de Bachillerato.
- El análisis estadístico de estas variables en los problemas planteados a lo largo de 12 años informa de las tendencias en las pruebas de acceso y puede servir a los diseñadores de las pruebas para mejorarla.
- La identificación y clasificación de los conflictos semióticos que muestran los participantes en el estudio respecto a la inferencia estadística puede servir para mejorar la formación de los futuros profesores, organizando acciones educativas que les permitan superarlos.
- Igualmente, la identificación de carencias en su formación didáctica es una información de utilidad a los formadores de profesores.
- Finalmente, resaltamos como otra aportación la síntesis del estado de la cuestión y las diferentes publicaciones surgidas de este trabajo.

5.5. LIMITACIONES DEL ESTUDIO Y FUTURAS LÍNEAS DE INVESTIGACIÓN

La principal limitación que se identifica de la investigación realizada sobre las pruebas de acceso a la universidad es la dificultad que hay para poder acceder a las respuestas de los estudiantes ante estos problemas. Sería de gran interés analizar dichas respuestas para poder realizar una mejor clasificación de las dificultades que encuentran los estudiantes de Bachillerato cuando se enfrentan a un problema de inferencia estadística y conseguir solventar dichas dificultades mediante una enseñanza adecuada.

Por otro lado, nos hemos limitado a los ejercicios propuestos en Andalucía. El estudio se podría ampliar comparando con los ejercicios de inferencia estadística propuestos en las pruebas de acceso a la universidad de las distintas comunidades, lo que nos podría permitir realizar un análisis comparativo y examinar cómo se trata en todo el territorio español este tema.

Igualmente se podría llevar a cabo un estudio de la presentación de la inferencia en los libros de texto de Bachillerato de Ciencias Sociales para comparar los ejercicios propuestos en los libros de texto y los utilizados en las PAU, y realizar recomendaciones para mejorar los textos, en caso de ser necesario.

Respecto a la evaluación con futuros profesores, una primera limitación es el tamaño de la muestra y el estar localizada en una única ciudad. Se podría continuar esta evaluación tomando otras muestras de futuros profesores en diferentes provincias españolas, para ver si los resultados se pueden generalizar a otros contextos.

Por otro lado, apenas se ha realizado evaluación de las componentes del conocimiento didáctico de los futuros profesores o de sus competencias de análisis didáctico. Esta sería otra línea de investigación relevante, para lo cual habría de organizarse nuevos talleres en que los futuros profesores realizasen ellos mismo el análisis semiótico de los problemas de inferencia, o analizaran la idoneidad didáctica de lecciones sobre inferencia o de pruebas de evaluación.

Queda también por analizar las actitudes de los profesores hacia el tema, puesto que sus actitudes influyen en las que desarrollan los estudiantes (Estrada, Batanero y Lancáster, 2011).

5.6. IMPLICACIONES PARA LA ENSEÑANZA

La importancia que dentro del aula se da a cada tema recogido en las orientaciones curriculares, en particular, para los estudiantes de segundo de Bachillerato viene implícitamente determinada por las pruebas de acceso a la universidad. Aunque la finalidad principal de las mencionadas pruebas es evaluar los conocimientos y capacidades adquiridas por los futuros universitarios durante su etapa de Bachillerato (primer y segundo curso), también se utilizan actualmente para seleccionar a los estudiantes que quieren ingresar en titulaciones y centros de estudios universitarios determinados.

Con el fin de garantizar el éxito de los estudiantes que se preparan para acceder a la universidad es vital asegurar una correcta y estrecha relación entre los contenidos de inferencia incluidos en las pruebas y los recogidos en los documentos curriculares. Los resultados obtenidos del análisis realizado en la investigación pueden servir para la

elaboración de pruebas futuras (incluida la nueva reválida) y preparar a los estudiantes que tienen que enfrentarse a las mismas. Además, esperamos que la información recogida en esta memoria contribuya a mejorar la práctica docente de los profesores de la etapa de Bachillerato.

También es fundamental para conseguir una buena enseñanza la formación de los profesores. En el taller donde se recogieron los datos de la evaluación se continuó en las sesiones posteriores con algunas actividades de simulación y de análisis de las respuestas dadas por los participantes. Pero el tiempo disponible fue claramente insuficiente para mejorar la formación de estos futuros profesores. Sería entonces necesario diseñar y experimentar, en base al tiempo disponible, tareas formativas incorporando, por ejemplo, aproximaciones informales a la inferencia estadística (Dolor y Noll, 2015) que traten de mejorar su formación y que refuercen sus conocimientos con objeto de solventar los conflictos semióticos detectados en la memoria de tesis presentada.

REFERENCIAS

- Alvarado, H. (2007). *Significados del teorema central del límite en la enseñanza de la estadística en ingeniería*. Tesis Doctoral. Universidad de Granada.
- Alvarado, H. y Batanero, C. (2007). Dificultades de comprensión de la aproximación normal a la distribución binomial. *Números*, 57. Disponible en: <http://www.sinewton.org/numeros>.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Autor.
- Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores*. Tesis Doctoral. Universidad de Granada.
- Artigue, M., Batanero, C. y Kent, P. (2007). Mathematics thinking and learning at post-secondary level. En F. Lestee (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1011-1049). Charlotte, NC: Information Age Publishing.
- Ball, D. L., Thames, M. H. y Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, 59(5), 389-407.
doi: 10.1177/0022487108324554
- Bakker, A. y Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical thinking and learning*, 13(1-2), 5-26.
doi: 10.1080/10986065.2011.538293
- Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning*, 2(1-2), 75-98.
doi: 10.1207/S15327833MTL0202_4
- Batanero, C. (2001). *Didáctica de la estadística*. Granada: Universidad de Granada.
- Batanero, C. (2020). Treinta años de investigación didáctica sobre el análisis inferencial de datos. En A. Ávila (Ed.), *Rutas de la educación matemática* (pp. 186-199). México, D.F.: Sociedad Mexicana de Investigación y Divulgación de la Educación Matemática-
- Batanero, C., Arteaga, P. y Gea, M. M. (2012). El currículo de estadística: reflexiones desde una perspectiva internacional, *UNO*, 59, 9-17.
- Batanero, C., Begué, N., Borovcnik, M. y Gea, M. M. (en prensa). Ways in which high-school students understand the sampling distribution for proportions. *Statistics Education Research Journal*.

- Batanero, C. y Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense publishers. doi: 10.1007/978-94-6300-624-8
- Batanero, C., Burrill, G. y Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study*. New York: Springer.
- Batanero, C. y Díaz-Batanero, C. (2006). Methodological and didactical controversies around statistical inference: *Actes des 38-ièmes Journées de Statistique*. [CD-ROM] París: Société Française de Statistique.
- Batanero, C., Díaz-Batanero, C. y López-Martín, M. M. (2017). Significados del contraste de hipótesis, configuraciones epistémicas asociadas y algunos conflictos semióticos. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Disponible en: enfoqueontosemiotico.ugr.es/civeos.html
- Batanero, C., Díaz-Batanero, C., López-Martín y Roldán, A. (2020). Estimación por intervalos, aproximaciones metodológicas y dificultades de comprensión. *Boletín de Estadística e Investigación Operativa*, 36(3), 1-23.
- Batanero, C., Gea, M. M. y Begué, N. (2019). El sentido del muestreo. *Números*, 100, 121-124.
- Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R. y Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547. doi: 10.1080/0020739940250406
- Batanero, C., Godino, J. D. y Roa, R. (2004). Training teachers to teach probability. *Journal of statistics Education*, 12(1). doi: 10.1080/10691898.2004.11910715
- Batanero, C. y López-Martín, M. M. (en prensa). Conocimiento del intervalo de confianza por futuros profesores de bachillerato. *Jornal Internacional de Estudos em Educação Matemática*.
- Batanero C., López-Martín M. M., Arteaga P. y Gea M. M. (2018). Characterizing probability problems posed in university entrance tests in Andalucía. En C. Batanero y E. Chernoff (Eds.) *Teaching and learning stochastic. ICME-13 monographs*. Cham: Springer. doi: 10.1007/978-3-319-72871-1_7
- Batanero, C., López-Martín, M. M., Gea, M. M. y Arteaga, P. (2018). Conocimiento del contraste de hipótesis por futuros profesores de educación secundaria y bachillerato. *Publicaciones*, 48(2), 73-95. doi:10.30827/publicaciones.v48i2.8334.
- Batanero, C., Tauber, L. y Sánchez, B. (2001). Significado y comprensión de la distribución normal en un curso de análisis de datos. *Quadrante*, 10(1), 59-92.
- Batanero, C., Vera, O. D. y Díaz-Batanero, C. (2012). Dificultades de estudiantes de Psicología en la comprensión del contraste de hipótesis. *Números*, 80, 91-101.
- Beckman, M. D., DelMas, R. C. y Garfield, J. (2017). Cognitive Transfer Outcomes for a Simulation-Based Introductory Statistics Curriculum. *Statistics Education Research Journal*, 16, 419-440. arXiv:1707.06537v1

- Begué, N., Batanero, C. y Gea, M. M. (2018). Comprensión del valor esperado y variabilidad de la proporción muestral por estudiantes de educación secundaria obligatoria. *Enseñanza de las Ciencias*, 36(2), 63-79. ISSN: 0212-4521. doi: 10.5565/rev/ensciencias.2256.
- 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. doi: 10.1037/1082-989X.10.4.389.
- Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics*, 4, 24-27. doi: 10.1111/j.1467-9639.1982.tb00451.x
- Bisquerra, R. y Alzina, R. B. (2004). *Metodología de la investigación educativa*. Madrid: Editorial La Muralla.
- Black, M. (1979). *Inducción y probabilidad*. Madrid: Cátedra.
- Blázquez, F. y Luengo, R. (1989). *Las calificaciones en las pruebas de acceso en la Universidad de Extremadura*. Badajoz: ICE, Universidad
- Boal, N., Bueno, C., Leris, M. D. y Sein-Echaluce, M. L. (2008). Las habilidades matemáticas evaluadas en las Pruebas de Acceso a la Universidad. Un estudio en varias universidades públicas españolas. *Revista de Investigación Educativa*, 26(1), 11-23.
- 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
- Burgos, M. J. (2020). *Niveles de algebrización en el razonamiento proporcional desde las perspectivas institucional y personal. Implicaciones para la formación de profesores de matemáticas*. Tesis Doctoral. Universidad de Granada.
- Burgos, M. y Godino, J. D. (2019). Conflictos semióticos de alumnos de primaria en la resolución de una tarea de porcentajes. En J. Marbán, M. Arce, A. Maroto, J. M. Muñoz-Escolano y A. Alsina (Eds.), *Investigación en Educación Matemática XXIII* (pp. 223-232). Valladolid, España: Universidad de Valladolid.
- Cabriá, S. (1994). *Filosofía de la estadística*. Valencia: Servicio de Publicaciones de la Universidad.
- Cañadas, G., Batanero, C., Díaz-Batanero, C. y Roa, R. (2012). Psychology students' understanding of the chi-squared test. *Statistique et Enseignement*, 3(1), 3-18.
- Caperos, J. M. y Pardo, A. (2013). Consistency errors in p-values reported in Spanish psychology journals. *Psicothema*, 25, 408-414. doi: 10.7334/psicothema2012.207.
- Caraballo, R. (2010). *Análisis de los ítems de las pruebas de evaluación de diagnóstico en competencia matemática para el segundo curso de la Educación Secundaria Obligatoria en España, 2008-2009: un estudio exploratorio*. Trabajo fin de Máster. Universidad de Granada.

- Carretero, M. (2014). *Análisis de problemas de probabilidad en las pruebas de acceso de Matemáticas Aplicadas a las Ciencias Sociales de Andalucía*. Trabajo fin de Máster. Universidad de Granada.
- Carretero, M., López-Martín, M. M. y Arteaga, P. (2015). Contenido matemático de los problemas de probabilidad en las pruebas de acceso de Andalucía. *Probabilidad Condicionada. Actas de las II Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria*. Granada: SEIEM. Disponible en: <http://www.jvdiesproyco.es/>,
- Castro-Sotos, A. E., Vanhoof, S., Noortgate, W. y Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98-113. doi: 10.1016/j.edurev.2007.04.001
- Castellanos, M. T. (2013). *Tablas y gráficos estadísticos en las pruebas Saber Colombia*. Trabajo fin de Máster. Universidad de Granada.
- 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.
- Chance, B., delMas, R. C. y Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Amsterdam: Kluwer. doi: 10.1007/1-4020-2278-6_13.
- Chaput, B., Girard, J. C., y Henry, M. (2011). Frequentist approach: modelling and simulation in statistics and probability teaching. En C. Batanero, G. Burrill, y C. Reading (Eds.), *Teaching Statistics in school mathematics-Challenges for teaching and teacher education* (pp. 85-95). New York: Springer. doi: 10.1007/978-94-007-1131-0_12.
- Colera, J., García, R. y Oliveira, M. J. (2003). *Matemáticas aplicadas a las Ciencias Sociales II*. Madrid: Anaya.
- Contreras, J. M., López-Martín, M. M., Arteaga, P. y Carretero, M. (2015, Febrero). Probability content in the entrance to university tests in Andalusia. Presentado en *International Conference. Turning data into knowledge: new opportunities for Statistics Education*. Lisboa: Instituto de Educação.
- Contreras, A., Ordoñez, L. y Wilhelmi, M. R. (2010). Influencia de las Pruebas de Acceso a la Universidad en la enseñanza de la integral definida en el Bachillerato. *Enseñanza de las Ciencias*, 28(3), 367-384.
- Cordero, F. y Flores, R. (2007). El uso de las gráficas en el discurso matemático escolar. Un estudio socioepistemológico en el nivel básico a través de los libros de texto. *Revista Latinoamericana de Matemática Educativa*, 10(1), 7-38.
- Cruz Sánchez, M. M. (2018). *Las matemáticas de la selectividad*. Trabajo fin de Máster. Universidad de Almería.
- Cumming, G. y Fidler, F. (2005, Agosto). Interval estimates for statistical communication: problems and possible solutions. Trabajo presentado en *IASE Satellite Conference on Communication of Statistics*. Sydney: International Association for Statistical Education.

- Cumming, G., Fidler, F., Kalinowski, P. y Lai, J. (2012). The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, 64(3), 138-146. doi: 10.1111/j.1742-9536.2011.00037.x.
- Cumming, G. y Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist*, 60, 170-180. doi: 0.1037/0003-066X.60.2.170.
- Cumming, G., Williams, J. y Fidler, F. (2004). Replication and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics*, 3(4), 299-311. doi: 10.1207/s15328031us0304_5.
- Cuxart, A. (2000). Modelos estadísticos y evaluación: tres estudios en educación. *Revista de Educación*, 323, 369-394
- 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.
- Díaz-Batanero, C. (2005). *Apuntes sobre inferencia bayesiana*. Granada: La autora.
- Díaz-Batanero, C. (2018). *La lógica del contraste de hipótesis*. Material para el concurso a profesor titular de universidad. Huelva: La autora.
- Díaz-Batanero, C., Contreras, J. M. Batanero, C. y Roa, R. (2012). Evaluación de sesgos en el razonamiento sobre probabilidad condicional en futuros profesores de educación Secundaria. *Bolema*, 26(22), 1207-1226. doi: 10.1590/S0103-636X2012000400006.
- Díaz-Batanero, C., Lozano-Rojas, O. M. y 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.
- Díaz Barriga, A. y Luna Miranda, A. B. (2014). *Metodología de la investigación educativa: Aproximaciones para comprender sus estrategias*. Ediciones Díaz de Santos.
- Díaz, P., Mier, V., Alonso P. y Rodríguez-Muñiz, L. J. (2014). Probability and statistics in access exams to Spanish universities. En K. Makar, B. de Sousa y R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, Julio, 2014)*, Flagstaff, AR: International Statistical Institute.
- Dolor, J. y Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal*, 14(1), 60-89.
- Espinel, M. C., Ramos, C. E. y Ramos. R. M. (2006, Septiembre). La inferencia estadística en la PAU. Presentado en *XXIX Congreso Nacional de Estadística e Investigación Operativa, SEIO*. Tenerife.
- Espinel, M. C., Ramos, R. M. y Ramos, C. E. (2007). Algunas alternativas para la mejora de la enseñanza de la inferencia estadística en Secundaria. *Números*, 67, 15-23.

- Estrada, A., Batanero, C. y Lancaster, S. (2011). Teachers' attitudes towards statistics. En *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 163-174). Springer, Dordrecht.
- Falk, R. (1986). Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292 – 297). Victoria, Canada: International Statistical Institute.
- Falk, R. y Greenbaum, C. W. (1995) Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology*, 5(1), 75-98. doi: 10.1177/0959354395051004.
- Fan, L. (2013). Textbooks research as scientific research: towards a common ground on issues and methods of research on mathematics textbooks. *ZDM*, 45(5), 765-777. doi: 10.1007/s11858-013-0530-6
- Fan, L., Zhu, Y. y Miao, Z. (2013). Textbook research in mathematics education: development status and directions. *ZDM*, 45(5), 633-646. doi: 10.1007/s11858-013-0539-x
- Fidler, F. (2005). *From statistical significance to effect estimation*. Tesis Doctoral. Universidad de Melbourne.
- Fidler, F. y Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th Session of the International Statistical Institute* (pp. 1-5). Voorburg: International Statistical Institute. Disponible en iase-web.org/documents/papers/isi55/Fidler-Cumming.pdf.
- Fisher, R. A. (1935). *The design of experiments*. Edimburgh: Oliver y Boyd.
- Font, V. y Godino, J. D. (2006). La noción de configuración epistémica como herramienta de análisis de textos matemáticos: su uso en la formación de profesores. *Educação Matemática Pesquisa*, 8(1), 67-98.
- Font, V., Godino, J. D. y Gallardo, J. (2013). The emergence of objects from mathematical practices. *Educational Studies in Mathematics*, 82, 97–124. doi: 10.1007/s10649-012-9411-0.
- Frías, M. D., Pascual, J. y García, J. F. (2002). La hipótesis nula y la significación práctica. *Metodología de las Ciencias del Comportamiento*, 4(especial), 181-185.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25. doi: 10.1111/j.1751-5823.2002.tb00336.x
- García, I. (2011). Análisis de los términos de inferencia estadística en Bachillerato. *Números*, 77, 51-73.
- García, I. y García, J.A. (2005). Algunos resultados sobre la actuación de los alumnos en las cuestiones de estadística en la P. A. U. *Actas de las XI Jornadas para el Aprendizaje y Enseñanza de las Matemáticas* (pp. 733-738). La Laguna: Consejería de Educación, Cultura de Deportes.
- García, I. y García, J. A. (2009). Enseñanza de la estadística y lenguaje: un estudio en Bachillerato. *Educación Matemática*, 21(3), 95-126. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-58262009000300005&lng=es&nrm=iso

- Garfield, J. y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer Science & Business Media.
- Garfield, J., delMas, R. y Chance, B. (2004). *Tools for teaching and assessing statistical inference*. Disponible en: http://www.gen.umn.edu/research/stat_tools.
- Gaviria, J. L. (2005). La equiparación del expediente de Bachillerato en el proceso de selección de alumnos para el acceso a la universidad. *Revista de Educación*, 337, 351-387.
- Gea, M. M. (2014). *La correlación y regresión en bachillerato: análisis de libros de texto y del conocimiento de los futuros profesores*. Tesis Doctoral. Universidad de Granada.
- Gigerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. En G. Keren y C. Lewis (Eds.), *A handbook for data analysis in the behavioural sciences: Methodological issues* (pp. 311 – 339). Hillsdale, NJ: Erlbaum.2003M
- Godino, J. D. (1996). Mathematical concepts, their meanings and understanding. En L. Puig y A. Gutiérrez (Eds.), *Proceedings of the 20th PME Conference* (Vol. 2, pp. 417-424). 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. (2003). *Teoría de las funciones semióticas: Un enfoque ontológico-semiótico de la cognición e instrucción matemática*. Granada: Servicio de reprografía de la Facultad de Ciencias, 2003. Disponible en: <http://www.ugr.es/~jgodino/funciones-semioticas/monografiatfs.pdf>
- Godino, J. D. (2009). Categorías de análisis de los conocimientos del profesor de matemáticas. *UNIÓN, Revista Iberoamericana de Educación Matemática*, 20, 13-31
- Godino, J. D. (2011, Julio). Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas. Trabajo presentado en la *XIII CIAEM-IACME*, Recife, Brasil.
- Godino, J. D. (2013). Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas. *Cuadernos de Investigación y Formación en Educación Matemática*, 11(1), 111-132.
- Godino, J. D. (2017). Construyendo un sistema modular e inclusivo de herramientas teóricas para la educación matemática. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Granada: Grupo FQM12. Disponible en: enfoqueontosemiotico.ugr.es/civeos.html
- Godino, J. D., y Batanero, C. (1994). Significado personal e institucional de los objetos matemáticos. *Recherches en Didactiques des Mathématiques*, 14(3), 325-355.
- Godino, J. D. y Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education. En A. Sierpinska y J. Kilpatrick (Eds.), *Mathematics education a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

- 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), 27-135. doi: 10.1007/s11858-006-0004-1.
- Godino, J. D., Batanero, C. y Font, V. (2020). El Enfoque ontosemiótico: implicaciones sobre el carácter prescriptivo de la didáctica. *Revista Chilena de Educación Matemática*, 12(2), 47-59. doi: 10.46219/rechiem.v12i2.25
- Godino, J. D., Batanero, C., Font, V. y Giacomone, B. (2016). *Articulando conocimientos y competencias del profesor de matemáticas: el modelo CCDM*. En J. A. Macías, A. Jiménez, J. L. González, M. T. Sánchez, P. Hernández, C. Fernández, F. J. Ruiz, T. Fernández y A. Berciano (Eds.), *Investigación en Educación Matemática XX*, (pp. 285-294). Málaga: SEIEM. (pp. 285-294). Malaga, España: Universidad de Málaga.
- Godino, J. D., Batanero, C., Rivas, H. y Arteaga, P. (2013). Componentes e indicadores de idoneidad de programas de formación de profesores en didáctica de las matemáticas. *Revemat*, 8(1), 46-74. doi: 10.5007/1981-1322.2013v8n1p46
- Godino, J. D., Batanero, C., Roa, R. y Wilhelmi, M. R. (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. Trabajo presentado en el *Joint ICMI / IASE Study Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education*. Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education.
- Godino, J.D., Bencomo, D., Font, V., y Wilhelmi, M. R. (2006) Análisis y valoración de la idoneidad didáctica de procesos de estudio de las matemáticas. *Paradigma*, 27(2), 221-252.
- Godino, J. D. y Burgos, M. (2019, Diciembre). Interweaving Transmission and Inquiry in Mathematics and Sciences Instruction. En *International Congress on Education and Technology in Sciences* (pp. 6-21). Springer, Cham.
- Godino, J. D. y Burgos, M. (2020). ¿Cómo enseñar las matemáticas y ciencias experimentales? Resolviendo el dilema entre transmisión e indagación. *Paradigma*, 41, 80-106.
- Godino, J. D., Burgos, M. y Wilhelmi, M. R. (2020). Papel de las situaciones adidácticas en el aprendizaje matemático: una mirada crítica desde el enfoque ontosemiótico. *Enseñanza de las Ciencias*, 38(1), 0147-164. doi: 10.5565/rev/ensciencias.2906
- Godino, J. D., Contreras, Á. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 26(76), 39-88.
- Godino, J. D., Font, V., Contreras, Á. y Wilhelmi, M. R. (2006). Una visión de la didáctica francesa desde el enfoque ontosemiótico de la cognición e instrucción matemática. *Revista latinoamericana de investigación en matemática educativa*, 9(1), 117-150. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-24362006000100006&lng=es&nrm=iso

- Godino, J. D., Giacomone, B., Batanero, C. y Font, V. (2017). Enfoque ontosemiótico de los conocimientos y competencias del profesor de matemáticas. *Boletim de Educação Matemática*, 31(57), 90-113. Disponible en: www.redalyc.org/articulo.oa?id=291250692007
- Godino, J. D., Ortiz, J. J., Roa, R. y Wilhelmi, M. R. (2011). Models for statistical pedagogical knowledge. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (Vol 14, pp. 271-282). Springer, Dordrecht. doi: 10.1007/978-94-007-1131-0_27Springer.
- Hernández, B., Rubén, M. González, C. y Martín, I. (2002). Perfeccionando la enseñanza de la inferencia estadística en el contexto de las ciencias médicas. *Revista Habanera de Ciencias Médicas*, 1(1), 51-61.
- Grau, R., Cuxart, A. y Martí-Recober, M. (2002). La calidad en el proceso de corrección de las pruebas de acceso a la Universidad: variabilidad y factores. *Revista de Investigación Educativa*, 20(1), 209-223.
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427-437. doi: 10.2307/30034960
- Hald, A. (2008). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. New York: Springer.
- Hacking, I. (2006). *The emergence of probability. A philosophical study of early ideas about probability, induction and statistical inference*. New York: Cambridge University Press.
- Harlow, L. L., Mulaik, S. A. y Steiger, J. H. (2016). *What if there were no significance tests?* Mahwah, NJ: Erlbaum. 2ª edición.
- 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. doi: 10.1007/978-94-007-1131-0_24.
- Hernández, C. E. y Carpio, N. (2019). Introducción a los tipos de muestreo. *Alerta, Revista científica del Instituto Nacional de Salud*, 2(1), 75-79. doi: 10.5377/alerta.v2i1.7535
- Hernández, R., Fernández, C. y Baptista, P. (1994). *Metodología de la investigación*. McGraw-Hill.
- Hill, H. C., Ball, D. L. y Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39, 372-400. Disponible en: <http://www.jstor.org/stable/40539304>
- Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, 8(1), 3 - 7. doi: 10.1111/j.1467-9280.1997.tb00534.x
- Inzunza, S. y Jiménez, J. V. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista Latinoamericana de Investigación en Matemática Educativa*, 16(2), 179-211. doi:

10.12802/relime.13.1622

- Krauss, S. y Wassner, C. (2002). How significance tests should be presented to avoid the typical misinterpretations. En C. Reading (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education. Disponible en: www.stat.auckland.ac.nz/~iase/publications.
- Krippendorff, K. (2013). *Content analysis: an introduction to its methodology*. London, Sage
- Lecoutre, B., Lecoutre M. P., y Poitevineau J. (2001). Uses, abuses and misuses of significance tests in the scientific community: Won't the Bayesian choice be unavoidable? *International Statistical Review*, 69, 399-418. doi: 10.1111/j.1751-5823.2001.tb00466.x
- Lipson, K. (1997). What do students gain from computer simulation exercises? En J. B. Garfield. y G. Burril (Eds.), *Proceedings of the IASE 1996 Round Table Conference*. (pp. 131-144). Vooburg, Holanda: International Statistical Institute e International Association for Statistical Education.
- Liu, Y. y Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies*, 4(2), 126-138. doi: 10.1080/15544800902741564
- Llinares, S. y Krainer, K. (2006). Mathematics (student) teachers and teacher educators as learners. En A. Gutierrez y P. Boero (Eds.) *Handbook of research on the psychology of mathematics education. Past, Present and Future* (pp. 429-459). Rotterdam /Taipei: Sense Publishers. doi: 10.1163/9789087901127_016
- López-Martín, M. M. (2016). *Variables que determinan los problemas de inferencia propuestos en Andalucía en las pruebas de acceso a la universidad*. Trabajo fin de Máster. Universidad de Granada.
- López-Martín, M. M., Batanero, C., Díaz-Batanero, C., y Gea, M. M. (2016). La inferencia estadística en las pruebas de acceso a la universidad en Andalucía. *Revista Paranaense de Educação Matemática*, 5(8), 33-59.
- López-Martín, M. M., Batanero, C., Gea, M. M. (2018). La faceta cognitiva en el conocimiento de futuros profesores sobre el contraste de hipótesis. En L. J. Rodríguez-Muñiz, L. Muñoz-Rodríguez, A. Aguilar-González, P. Alonso, F. J. García García y A. Bruno (Eds.), *Investigación en Educación Matemática XXII* (pp. 300-309). Gijón: SEIEM.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística? *Bolema. Boletim de Educação Matemática*, 33(64), 672-693. doi: 10.1590/1980-4415v33n64a11
- López-Martín, M. M., Batanero, C., Gea, M. y Arteaga, P. (2016), Análisis de los problemas de inferencia propuestos en las Pruebas de Acceso a la Universidad en Andalucía, *Vidya* 36(2), 409-428.
- López-Martín, M. M., Contreras, J. M., Batanero, C. y Carretero, M. (2015). Los problemas de probabilidad propuestos en las Pruebas de Acceso a la Universidad en Andalucía. *Areté*, 1(1), 39-60.

- López-Martín, M. M., Contreras, J. M., Carretero, M., Serrano, L. (2016). Análisis de los problemas de probabilidad propuestos en las Pruebas de Acceso a la Universidad en Andalucía. *Avances de Investigación en Educación Matemática*, 9, 65-84. doi: 10.35763/aiem.v0i9.109
- López-Martín, M. M., Molina-Portillo, E., Contreras, J. y Ruz, F. (2019). Análisis de los errores inferenciales en el ámbito científico. 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.
- Mallart, A. (2014). La resolución de problemas en la prueba de Matemáticas de acceso a la universidad: procesos y errores. *Educatio Siglo XXI*, 32(1), 233-254. doi: 10.6018/j/194171.
- Mayo, D. G. y Cox, D. R. (2006). Frequentist statistics as a theory of inductive inference. *IMS Lecture Notes-Monograph Series*, 49, 77-97. doi:10.1214/074921706000000400
- MacGillivray, H. y Pereira Mendoza, L. (2011). Teaching statistical thinking through investigative projects. En C. Batanero, G. Burrill, y C. Reading, (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI and IASE study* (pp. 109-120). New York: Springer.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Sussex, UK: Psychology Press.
- Mendonça, T., Coutinho, C. y Almouloud, S. (2006). Mathematics education and statistics education: meeting points and perspectives. En *Proceedings of the Seventh International Conference on Teaching Statistics*. CD ROM. Salvador (Bahia), Brazil: International Association for Statistical Education and International Statistical Institute.
- MCD, Ministerio de Educación, Cultura y Deporte (2013). *PISA 2012. Programa para la evaluación internacional de los estudiantes. Informe español. Vol.1. Resultados y contexto*. Madrid: Autor.
- ME, Ministerio de Educación (2009). *PISA 2009. Programa para la evaluación internacional de estudiantes de la OCDE. Informe español*. Madrid: Autor.
- MEC, Ministerio de Educación y Ciencia (2007). *Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del Bachillerato y se fijan sus enseñanzas mínimas*. Madrid: Autor.
- MECD, Ministerio de Educación, Cultura y Deporte (1999). *Real Decreto 1640/1999, de 22 de octubre, por el que se regula la prueba de acceso a estudios universitarios*. Madrid: Autor.
- MECD, Ministerio de Educación, Cultura y Deporte (2000). *Real Decreto 990/2000, de 2 de junio, por el que se modifica y completa el Real Decreto 1640/1999, de 22 de octubre, por el que se regula la prueba de acceso a estudios universitarios*. Madrid: Autor.
- MECD, Ministerio de Educación, Cultura y Deporte (2002). *Real Decreto 1025/2002, de 4 de octubre, por el que se modifica el Real Decreto 1640/1999, de 22 de octubre, modificado y completado por el Real Decreto 990/2000, de 2 de junio, por el que se regula la prueba de acceso a estudios universitarios*. Madrid: Autor.

- MECD, Ministerio de Educación, Cultura y Deporte (2003). *Real Decreto 1741/2003, de 19 de diciembre, por el que se regula la prueba general de Bachillerato*. Madrid: Autor.
- MECD, Ministerio de Educación, Cultura y Deporte (2014). *Real Decreto 412/2014, de 6 de junio, por el que se establece la normativa básica de los procedimientos de admisión a las enseñanzas universitarias oficiales de Grado*. Madrid: Autor.
- MECD, Ministerio de Educación, Cultura y Deporte (2015a). *Proyecto de real decreto por el que se regulan las características generales de las pruebas de la evaluación final de Educación Primaria, y las características de las pruebas de las evaluaciones finales de Educación Secundaria Obligatoria y de Bachillerato, establecidas en la Ley Orgánica 2/2006, de 3 de mayo, de Educación*. Madrid: Autor.
- MECD, Ministerio de Educación, Cultura y Deporte (2015b). *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.
- Méndez, H. (1991). *Understanding the central limit theorem*. Tesis Doctoral. Universidad de California.
- Mengual, E., Gorgorió, N. y Albarracín, L. (2013). Validación de un instrumento para la calificación de exámenes de matemáticas. En A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 367-381). Bilbao: SEIEM.
- Mingorance, C. (2014). *La estadística en las pruebas de diagnóstico andaluzas*. Trabajo fin de Grado. Universidad de Granada.
- Moore, D. S. (2010). *The basic practice of statistics*. New York: Freeman (5th edition).
- MP, Ministerio de la Presidencia (2008). *Real Decreto 1892/2008, de 14 de noviembre, por el que se regula las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades públicas españolas*. Madrid: Autor.
- Moses, L. E. (1992). The reasoning of statistical inference. En D. C. Hoaglin y D. S. Moore (Eds.), *Perspectives on contemporary statistics* (pp. 107-122). Washington, DC: Mathematical Association of America.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D. y Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103-123.
- Mullis, I. V., Martin, M. O., Foy, P. y Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Muñoz Vitoria F. (1995). El acceso a la Universidad en España: perspectiva histórica. *Revista de Educación*, (308), 31-61. Disponible en: <http://hdl.handle.net/11162/70534>
- Murillo, F. J. (1997). Análisis de las pruebas que conforman la selectividad. *Revista de Educación*, 314, 49-62.
- Neuendorf, K. A. (2016). *The content analysis guidebook*. London: Sage.

- Neyman J. y Pearson E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A*, 231, 289-337. doi:10.1098/rsta.1933.0009
- Nortes, A. N. y Nortes, R. (2010). Resolución de problemas de matemáticas en las Pruebas de Acceso a la Universidad. Errores significativos. *Educatio Siglo XXI*, 28(1), 317-342. Disponible en: <https://revistas.um.es/educatio/article/view/109851>
- Nortes, A. N., Nortes, R. y Lozano, F. (2015). Las correcciones en Matemáticas en las Pruebas de Acceso a la Universidad. *Educatio Siglo XXI*, 33(3), 199-222. Disponible en: <http://hdl.handle.net/10201/49484>
- OCDE, Organización para la Cooperación y el Desarrollo Económico (2015). *PISA 2015 mathematics framework*. Paris: OCDE
- Occelli, M. y Valeiras, B. N. (2013). Los libros de texto de ciencias como objeto de investigación: una revisión bibliográfica. *Enseñanza de las Ciencias*, 31(2), pp. 133-152. doi: 10.5565/rev/ec/v31n2.761
- Olivo, E. (2008). *Significados del intervalo de confianza en la enseñanza de la ingeniería en México*. Tesis Doctoral. Universidad de Granada.
- Ortiz, J. J., Mohamed, N., Serrano, L. y Albanese, V. (2017). La estimación de la media: análisis del lenguaje en libros de texto de Bachillerato. En J. M. Muñoz-Escolano, A. Arnal-Bailera, P. Beltrán-Pellicer, M. L. Callejo y J. Carrillo (Eds.), *Investigación en Educación Matemática XXI* (pp.377-386). Zaragoza: SEIEM
- Pino-Fan, L. R. y Godino, J. D. (2015). Perspectiva ampliada del conocimiento didáctico-matemático del profesor. *Paradigma*, 36(1), 87-109.
- Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Ramos, C. E., Espinel, M. C. y Ramos. R. M. (2009). Identificación de los errores en los contrastes de hipótesis de los alumnos de Bachillerato. *SUMA*, 61, 35-44.
- Reys, B. J., Reys, R. E. y Chavez, O. (2004). Why Mathematics Textbooks Matter. *Educational Leadership*, 61(5), 61-66.
- Rico, L. (2006a). Marco teórico de evaluación en PISA sobre matemáticas y resolución de problemas. *Revista de Educación*, (Extraordinario 2006), 275-294.
- Rico, L. (2006b). La competencia matemática en PISA. *PNA*, 1(2), 47-66.
- Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropos.
- Rodrigo, M. F., Molina, J. G., García-Ros, R. y Pérez-González, F. (2012). Efectos de interacción en la predicción del abandono en los estudios de Psicología. *Anales de Psicología/Annals of Psychology*, 28(1), 113-119.
- Rodríguez-Muñiz, L. J. y Díaz, P. (2018). Las investigaciones sobre la estadística y la probabilidad en los libros de texto de Bachillerato. ¿Qué se ha hecho y qué se puede hacer? *Avances de Investigación en Educación Matemática*, (14), 65-81.
- Roldán, A., Batanero, C. y Alvarez-Arroyo, R. (2020). Comprensión del intervalo de confianza por estudiantes de Bachillerato. *Avances de Investigación en Educación Matemática*, 18, 103-117. doi: 10.35763/aiem.v0i18.

- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.
- Ruiz, J., Dávila, P., Etxeberria, J. y Sarasua, J. (2013). Pruebas de selectividad en Matemáticas en la UPV-EHU. Resultados y opiniones de los profesores. *Revista de Educación*, 362, 217-246.
- Ruiz, J., Sarasua, J. y García, J. M. (2011). Una tipología y clasificación de los ejercicios de matemáticas de selectividad. *Epsilon*, 78, 21-38.
- Saldanha. L. y Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257-270.
- Saldanha. L. y Thompson, P. (2007) Exploring connections between sampling distributions and statistical inference: an analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, 2(3), 270-297. Disponible en: www.iejme.com/.
- Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere-Jones (Ed.), *Proceeding of the Third International Conference on Teaching Statistics* (pp. 486-490). Voorburg, The Netherlands: International Statistical Institute.
- Shadish, W. R., Cook, T. D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sharma, S. (2017) Definitions and models of statistical literacy: a literature review, *Open Review of Educational Research*, 4(1), 118-133. doi: 10.1080/23265507.2017.1354313.
- Shaughnessy, J. M. y Ciancetta, M. (2004). Types of Student Reasoning on Sampling Tasks. En M. J. Hoines and A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, (Vol 4 pp 177-184). Bergen: Bergen University College.
- Shulman (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of new reform. *Harvard Educational Review*, 57(1), 1-22.
- Stohl, H. (2005). Probability in teacher education and development. En G. Jones (Ed.), *Exploring probability in school* (pp. 345-366). Boston, MA: Springer,
- Tamayo, M. (2004). *El proceso de la investigación científica*. México: Limusa.
- Tobias-Lara, M. G. y Gomez-Blancarte, A. L. (2019). Assessment of informal and formal inferential reasoning: A critical research review. *Statistics Education Research Journal*, 18(1). 8-25.
- Valera, S., Sánchez, J. y Marín, F. (2000). Contraste de hipótesis e investigación psicológica española: Análisis y propuestas. *Psicothema*, 12(2), 549-582.
- Vallecillos, A. (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral. Universidad de Granada,
- Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing

- hypotheses. *Proceedings of the 52 session of the International Statistical Institute* (Vol.2, pp. 201–204). Helsinki: International Statistical Institute.
- Wild, C. J. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-248.
doi: 10.1111/j.1751-5823.1999.tb00442.x.
- Wilhelmi, M. R., Godino, J. D. y Lacasta, E (2007). Configuraciones epistémicas asociadas a la noción de igualdad de números reales. *Recherches en Didactique des Mathematiques*, 27(1), 77-120
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
doi: 10.1037/0003-066X.54.8.594.
- 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. <https://doi.org/10.4324/9780203767740>.
- Yáñez, G. y Behar, R. (2009). Interpretaciones erradas del nivel de confianza en los intervalos de confianza y algunas explicaciones plausibles. En M. J. González, M.T. González y J. Murillo (Eds.). *Investigación en Educación Matemática XIII, Comunicaciones de los grupos de trabajo*. Santander: SEIEM.

ANEXO. TRABAJOS QUE DESARROLLAN ESTA MEMORIA

Artículos

- Álvarez-Arroyo, R. Roldán, A. y López-Martín, M. M. (en prensa). Propuestas para la mejora de la interpretación del intervalo de confianza en estudiantes preuniversitarios y universitarios. *Números*.
- Batanero, C. y López-Martín, M. M. (en prensa). Conocimiento del intervalo de confianza por futuros profesores de bachillerato. *Jornal Internacional de Estudos en Educaçao Matemática*.
- Batanero, C., Díaz-Batanero, C., López-Martín, M. M. y Roldán, A. (2020). Estimación por intervalos, aproximaciones metodológicas y dificultades de comprensión. *Boletín de Estadística e Investigación Operativa*, 36 (3), 1-23.
- Batanero, C., López-Martín, M.M., Gea, M. y Arteaga, P. (2019). Conocimiento del contraste de hipótesis por futuros profesores de educación secundaria y bachillerato. *Publicaciones* 48(2), 73-95. doi: 10.30827/publicaciones.v48i2.8334
- 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 Educaçao Matemática* 5 (8), 33-59.
- López-Martín, M.M., Batanero, C. y Gea, M.M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística? *Bolema. Boletim de Educaçao Matemática*. 32(60), 134-155. doi: 10.1590/1980-4415v33n64a11.
- López-Martín, M. M., Batanero, C., Gea, M. y Arteaga, P. (2016), Análisis de los problemas de inferencia propuestos en las Pruebas de Acceso a la Universidad en Andalucía, *Vidya* 36(2), 409-428.

Capítulos de libros

- Batanero, C., López-Martín, M.M., Arteaga, P. y Gea, M. M. (2018). Characterizing the probability problems proposed at the university entrance tests in Andalucía. En C. Batanero y E. Chernoff (Eds.), *Teaching and learning stochastics: advances in probability education research* (pp. 103-124). Heidelberg: Springer.

Congresos

- Batanero, C., Díaz-Batanero, C. y López-Martín, M. M. (2017). Significados del contraste de hipótesis, configuraciones epistémicas asociadas y algunos conflictos

- semióticos. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Disponibles en enfoqueontosemiotico.ugr.es/civeos.html
- Batanero, C., López-Martín, M. M., Gea, M. M. y Arteaga, P. (2018). Knowledge of statistical tests by prospective high school teachers. E. Bergqvist, M. Österholm, C. Granberg y L. Sumpter (Eds.), *42nd Conference of the International Group for the Psychology of Mathematics Education*, (vol. 2, pp. 131-138). Umeå, Suecia. Julio, 2018.
- Batanero, C., López-Martín, M. M., González-Ruiz, I. y Díaz-Levicoy, D. (2015). Las medidas de dispersión en el estudio de la inferencia estadística. En M. Parraguez, H. Rivas, C. Vásquez, N. Pincheira, H. Solar, F. Rojas y E. Chandía (Eds.), *XIX Jornadas Nacionales de Educación Matemática*. Villarrica-Chile: Sociedad Chilena de Educación Matemática.
- López-Martín, M. M., Batanero, C., Contreras, J. M. (2017) El contraste de hipótesis en las pruebas andaluzas de acceso a la universidad. presentado en RELME 31, Lima. 2017. *Acta Latinoamericana de Matemática Educativa* 31, 931-938.
- López-Martín, M. M., Arteaga, P., Gea, M. M. y Contreras, J. M. (2016, Julio). Los intervalos de confianza en las pruebas de selectividad andaluzas. Trabajo presentado en la 30 *Reunión Latinoamericana de Matemática Educativa, RELME-30*, Monterrey, México, 2016. Comité Latinoamericano de Matemática Educativa.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2017). Prospective teachers' procedures in solving a statistical test problem. En L. Gómez Chova, A. López Martínez, I. Candel Torres (Eds.), *10th annual International Conference of Education, Research and Innovation*. International Academy of Technology, Education and Development (pp. 3058-3064). Sevilla: IATED Academy. ISSN: 2340-1095.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2018). La faceta cognitiva en el conocimiento de futuros profesores sobre el contraste de hipótesis. En L. J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, A. Aguilar-González, P. Alonso, F. J. García García y A. Bruno (Eds.), *Investigación en Educación Matemática XXII* (pp. 300-309). Gijón: SEIEM.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2019, Febrero). Prospective high school teachers' interpretation of hypothesis tests and confidence intervals. En Jankvist, U. T., Van den Heuvel-Panhuizen, M. y Veldhuis, M. (Eds.). (2019). *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education*. Utrecht, the Netherlands: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2018). La faceta cognitiva en el conocimiento de futuros profesores sobre el contraste de hipótesis. En L. J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, A. Aguilar-González, P. Alonso, F.G. García, y A. Bruno (Eds.), *Investigación en Educación Matemática XXII* (pp. 300-309). Gijón, España: Universidad de Oviedo.
- López-Martín, M. M., Batanero, C., Gea, M. M. y Arteaga, P. (2018). Assessing prospective high school teachers' understanding of statistical hypotheses. En L. Gómez Chova, A. López Martínez y I. Candel Torres (Eds.), *12th International*

Technology, Education and Development Conference (pp. 2808-2813). Valencia: IATED Academy.

- López-Martín, M. M., Batanero, C., Gea, M. M., González-Ruiz, I. (2015). Analysing the statistical inference problems proposed in the University entrance test in Andalucía. (2015). En L. Gómez, A. López e I. Candel (Eds.), *8th Annual International Conference of Education, Research and Innovation. International Academy of Technology, Education and Development (IATED)*. Sevilla: IATED Academy. Disponible en: library.iated.org/view/LOPEZMARTIN2015ANA
- López-Martín, M. M., Díaz-Levicoy, D., Gea, M. M. y Arteaga, P. (2016, Junio). Estudio empírico de los problemas sobre intervalos de confianza en las Pruebas de Acceso a la Universidad. Presentado en el *VI Encuentro Internacional en la Enseñanza de la Probabilidad y la Estadística (EIEPE)*. Ciudad de Puebla, Puebla, México, 2016.
- López-Martín, M.M., Molina-Portillo, E., Contreras, J. y Ruz, F. (2019). Análisis de los errores inferenciales en el ámbito científico. 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 .
- Roldán López de Hierro, A., López-Martín, M. M. y Batanero, C. (2019, Febrero). An alternative method to compute confidence intervals for proportion. En Jankvist, U. T., Van den Heuvel-Panhuizen, M. y Veldhuis, M. (Eds.). (2019). *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education*. Utrecht, the Netherlands: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.