

¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística?

Do prospective teachers know their students' errors in statistical inference?

Maria del Mar López-Martín*

 ORCID iD 0000-0001-8677-9606

Carmen Batanero**

 ORCID iD 0000-0002-4189-7139

María M. Gea***

 ORCID iD 0000-0002-5229-0121

Resumen

Se presenta un estudio de evaluación de los conocimientos de un grupo de futuros profesores españoles de Educación Secundaria y Bachillerato, en relación a los errores de sus estudiantes en la inferencia estadística. Dicho conocimiento forma parte de la faceta cognitiva del conocimiento-didáctico matemático del profesor. Para llevar a cabo la evaluación se pidió a setenta estudiantes, que se preparaban como futuros profesores, describir los errores previsibles de sus estudiantes al realizar un contraste de hipótesis y un intervalo de confianza, después de haber resuelto ellos mismos un problema de cada uno de estos tipos. Las respuestas obtenidas se clasifican en relación con diferentes etapas de los procedimientos (selección del procedimiento, planteamiento, errores conceptuales, procedimentales e interpretación) y las categorías en cada una de estas etapas se determinan teniendo en cuenta los errores descritos en la investigación sobre comprensión de la inferencia estadística. Los resultados muestran cierto conocimiento de los errores frecuentes por parte de los futuros profesores, aunque, en general, falta precisión y hay poca consciencia de los errores relacionados con el nivel de significación y p -valor.

Palabras clave: Inferencia estadística. Conocimiento didáctico-matemático. Futuros profesores.

Abstract

The aim of this work was to assess the knowledge of students' errors in statistical inference in a group of Spanish prospective High and Secondary School teachers. This knowledge is part of the cognitive facet of teachers' didactic-mathematical knowledge. A sample of seventy prospective teachers was asked to describe their students' most likely error in carrying out a statistical test and a confidence interval, after having solved themselves a problem of each type. The responses are classified in agreement with the different steps in the procedures (selecting a procedure, problem setting, conceptual, procedural, and interpretative errors). The categories in each

* Doctora en Técnicas Avanzadas en Gestión Empresarial. Universidad de Granada, España (UGR). Profesora Contratada Doctora de la Universidad de Almería (UAL), Almería, España. Universidad de Almería, Ctra. del Sacramento, s/n, 04120 La Cañada, Almería, España. E-mail: mdm.lopez@ual.es.

** Doctora en Ciencias Matemáticas, Universidad de Granada, España (UGR). Catedrática de Universidad de la Universidad de Granada (UGE), Granada, España. Facultad de Ciencias de la Educación, Campus de Cartuja, 18071 Granada, España. E-mail: batanero@ugr.es.

*** Doctora en Ciencias de la Educación, Universidad de Granada, España (UGR). Profesora Ayudante Doctor de la Universidad de la Universidad de Granada (UGE), Granada, España. Facultad de Ciencias de la Educación, Campus de Cartuja, 18071 Granada, España. E-mail: mmgea@ugr.es.

of these stages are determined by considering the errors described in research on understanding statistical inference. The results suggest a medium knowledge of the most frequent errors by prospective teachers; however, there is a general lack of precision and little awareness of errors linked to the significance level and p -value.

Keywords: Statistical inference. Didactic-mathematical knowledge. Prospective teachers.

1 Introducción

La inferencia estadística es ampliamente utilizada en la investigación y se ha incluido en España en Bachillerato (MEC, 2007; MECD, 2015) y en las pruebas de acceso a la universidad (LÓPEZ-MARTÍN; BATANERO; DÍAZ-BATANERO; GEA, 2016). Algunos autores como McLean (2002) consideran que unos conocimientos básicos de inferencia son parte de la cultura estadística de un ciudadano bien formado. Transmitir estos conocimientos a los estudiantes no es sencillo, pues diferentes revisiones de la literatura describen numerosos errores de comprensión y aplicación de estas técnicas (HARRADINE; BATANERO; ROSSMAN, 2011; MAKAR; RUBIN, 2018).

Es claro que una enseñanza efectiva de cualquier contenido requiere un conocimiento suficiente del mismo por parte del profesor y por ello, actualmente, encontramos una amplia investigación centrada en describir los componentes necesarios en el conocimiento del profesor de matemáticas (véase, por ejemplo, DAWSON; JAWORSKI; WOOD, 2013; KRAINER, 2014; LLINARES, 2018).

El presente trabajo se centra en evaluar la componente cognitiva del conocimiento del profesor, siguiendo el modelo de conocimiento didáctico-matemático (CDM) (GODINO, 2009; GODINO; BATANERO; FONT; GIACOMONE, 2016; PINO-FAN; GODINO, 2015). El modelo CDM considera que el conocimiento del profesor está formado por tres componentes principales: conocimiento matemático, conocimiento didáctico y conocimiento meta-matemático. La componente cognitiva es parte del conocimiento didáctico y se refiere al conocimiento que tiene el profesor de cómo los estudiantes aprenden, razonan y entienden las matemáticas, de sus estrategias de resolución de problemas y de los errores que cometen en el proceso de aprendizaje (PINO-FAN, GODINO, 2015).

El objetivo de este trabajo es evaluar una parte de esta componente; concretamente, el conocimiento que tienen los futuros profesores sobre los previsibles errores de sus estudiantes en dos procedimientos inferenciales: el contraste de hipótesis y el intervalo de confianza. A continuación, se describen los antecedentes, métodos y resultados del presente trabajo.

2 Antecedentes

Nos basamos en dos tipos de investigaciones previas; en primer lugar, las que se centran en el conocimiento de inferencia del profesor y, en segundo lugar, las que describen errores en estudiantes universitarios o de bachillerato, que compararemos con los previstos por los futuros profesores.

2.1 Conocimiento del profesor sobre inferencia

La escasa investigación previa sobre el conocimiento del profesor en inferencia se ha centrado, generalmente, en el conocimiento matemático (HARRADINE et al., 2011). Entre ellas encontramos la llevada a cabo por Haller y Krauss (2002), en la que se describen errores de profesores de cursos de metodología al interpretar el p -valor (probabilidad de encontrar un valor del estadístico muestral igual o más extremo que el observado, suponiendo la hipótesis nula cierta).

Por su parte, Liu y Thompson (2009) analizan las dificultades de ocho futuros profesores de matemáticas sobre inferencia, encontrando en ellos una concepción determinista del contraste de hipótesis y falta de comprensión de la lógica del mismo. Inzunsa y Jiménez (2013) caracterizan en cinco niveles el razonamiento de once futuros profesores de matemáticas sobre diversos conceptos ligados al contraste de hipótesis.

Los autores (INZUNSA; JIMÉNEZ, 2013) encontraron un alto porcentaje de estudiantes situados en el nivel más bajo de comprensión del nivel de significación, p -valor, relación con el tamaño de muestra, significado de una prueba de hipótesis e interpretación de resultados de una prueba de hipótesis. Asimismo, señalan que son pocos los estudiantes ubicados en los niveles superiores. El mejor desempeño se encontró en la formulación de hipótesis y en relacionar el p -valor con la significación estadística.

Estas dificultades no desaparecen con las nuevas propuestas de enseñanza de la inferencia con métodos informales basados en la formulación, pues en la experiencia de Biehler, Frischemeier y Podworny (2015) los profesores participantes continúan teniendo mucha dificultad con la formulación de hipótesis e interpretación del p -valor. No hemos encontrado antecedentes que describan el conocimiento matemático sobre el intervalo de confianza y el único que trata del conocimiento didáctico de los futuros profesores sobre el contraste de hipótesis es una versión preliminar y más limitada del trabajo que ahora presentamos (LOPEZ-MARTÍN; BATANERO; GEA, 2018).

2.2 Dificultades de los estudiantes con la inferencia

El principal error en el contraste de hipótesis consiste en intercambiar los términos de la probabilidad condicional en la definición del nivel de significación, interpretándolo como probabilidad de que la hipótesis nula sea cierta, una vez que se ha rechazado (VALLECILLOS, 1999).

Badenes-Ribera et al. (2015) y Caperos y Pardo (2013) encontraron el mismo error en la interpretación del p -valor. Otros errores ocurren cuando, por ejemplo, los estudiantes no identifican el tipo de contraste bilateral o unilateral (ESPINEL; RAMOS; RAMOS, 2007). Es también frecuente confundir la hipótesis nula con la alternativa, plantear hipótesis no complementarias o que no cubren el espacio paramétrico (VALLECILLOS, 1999; VERA; DÍAZ; BATANERO, 2011). Igualmente, se ha observado el uso de los estadísticos muestrales, en vez de los parámetros poblacionales en el establecimiento de las hipótesis (ESPINEL; RAMOS; RAMOS, 2007; HARRADINE et al., 2011; RAMOS; ESPINEL; RAMOS, 2009; VALLECILLOS, 1999; VERA et al., 2011).

Respecto al intervalo de confianza, Fidler y Cumming (2005) mostraron que los intervalos de confianza eran considerados como estadísticos descriptivos, ignorando su naturaleza inferencial. Otros estudios señalan la dificultad en percibir la relación entre amplitud del intervalo, tamaño de la muestra y coeficiente de confianza (BEHAR, 2001; CUMMING; WILLIAMS; FIDLER, 2004).

En un trabajo con estudiantes universitarios, de máster y futuros investigadores (HOEKSTRA et al., 2014), se observó en todos los grupos una alta proporción de sujetos que suponen que el nivel de confianza es la probabilidad de que la media muestral pertenezca al intervalo, o el porcentaje de muestras diferentes en que el intervalo coincidiría al repetir su cálculo.

Otros estudiantes confunden estadístico y parámetro o no recuerdan la expresión de la distribución muestral (BATANERO, 2000). A todo ello se ha de añadir la tipificación o uso incorrecto de la tabla de probabilidades de la distribución Normal e interpretar erróneamente los resultados (ESPINEL; RAMOS; RAMOS, 2007).

El estudio sistemático y clasificación de estos errores nos ha permitido elaborar una pauta para analizar el conocimiento de los futuros profesores de Educación Secundaria y Bachillerato sobre las dificultades a las que, previsiblemente, se enfrentarán sus estudiantes al trabajar con la inferencia. A continuación, presentamos el método, resultados y discusión de nuestra investigación.

Una versión piloto del mismo (LÓPEZ-MARTÍN; BATANERO; GEA, 2018) sirvió para realizar una identificación previa de las categorías utilizadas respecto al contraste de hipótesis. En el trabajo actual se revisan y extienden estas categorías, se completan con las correspondientes al intervalo de confianza y se desarrolla y extiende la comparación con las investigaciones previas.

3 Método

La muestra estuvo formada por setenta estudiantes del Máster de Formación del Profesorado de Educación Secundaria y Bachillerato, especialidad de matemáticas, que es obligatorio para aquellos que desean optar a una plaza de profesor. El 56% de los participantes eran licenciados en Matemáticas o Estadística, mientras que el resto eran ingenieros, arquitectos o científicos. Todos ellos habían estudiado la inferencia estadística y más de la mitad (57%) indicaron tener experiencia de enseñanza de las matemáticas.

La recogida de datos se realizó en un taller de inferencia dirigido a la formación matemática y didáctica de los futuros profesores, al que se dedicaron dos sesiones de dos horas de duración cada una. En la primera sesión, los participantes resolvieron dos problemas sobre inferencia estadística semejantes a los propuestos en años anteriores en las pruebas de acceso a la universidad (PAU) y que se presentan en el Cuadro 1.

Problema 1. La esperanza media de vida en un estudio desarrollado por las Naciones Unidas es de 69,2 años y su desviación típica 10. Se ha tomado una muestra 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: hipótesis planteadas, contraste elegido y su desarrollo e interpretación del resultado.

Problema 2. Determine un intervalo de confianza del 95% para la esperanza media de vida en los países europeos, explicando con detalle todos los pasos seguidos, ¿Cómo explicaría a un estudiante el significado de un intervalo de confianza y alguna de sus propiedades?

Cuadro 1 – Problemas resueltos por los participantes

Fuente: Elaborada por las autoras.

En el primer problema (contraste de hipótesis) se les pidió que planteasen las hipótesis, eligieran el contraste adecuado y lo desarrollaran, tomaran una decisión sobre el rechazo de la hipótesis nula e interpretasen los resultados en el contexto del problema. En el segundo problema (intervalo de confianza) se les pidió que determinasen un intervalo de confianza, solicitándoles, posteriormente, la interpretación y descripción de los factores de los cuáles depende la amplitud del intervalo. Para concluir la sesión se procedió a la corrección

de los problemas.

En la segunda sesión, los futuros profesores desarrollaron individualmente las tareas que analizamos en el presente trabajo, dirigidas a evaluar la faceta cognitiva de su conocimiento didáctico-matemático. Dichas tareas consistieron en responder por escrito, con el mayor detalle posible, a las siguientes preguntas:

Tarea 1. ¿Podrías indicar los errores que podrían cometer los estudiantes al realizar un contraste de hipótesis?

Tarea 2. ¿Podrías indicar los errores que podrían cometer los estudiantes al realizar un intervalo de confianza?

Se realizó un análisis de contenido a las respuestas proporcionadas por los futuros profesores. Este método, propio de la investigación cualitativa, permite el estudio sistemático y la realización de inferencias sobre el contenido de documentos escritos (NEUENDORF, 2016).

Mediante un procedimiento cíclico e iterativo se clasificaron los errores descritos por los estudiantes en los siguientes apartados: a) errores en la selección o planteamiento; b) errores conceptuales; c) errores procedimentales; y d) errores de interpretación del resultado de los cálculos matemáticos. En la primera tarea también se consideraron los errores citados respecto al planteamiento de las hipótesis. Dentro de cada uno de los apartados, se clasificaron las respuestas, teniendo en cuenta los errores típicos descritos en la investigación previa.

4 Resultados y discusión

Se analizaron un total de 491 errores descritos por los setenta futuros profesores (en promedio, 7,01 errores por cada uno). La Tabla 1 resume la frecuencia y porcentaje de errores citados en cada apartado del análisis y, la Tabla 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.

Los errores más citados en ambas tareas son los procedimentales, seguidos por los de interpretación de resultados. 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 1 - Número de errores citados y porcentaje (respecto al total de errores)

Categorías	Contraste		Intervalo	
	N. errores	%	N. errores	%
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

Fuente: elaborado por el autor.

Tabla 2 - Número de futuros profesores que citan algún error en cada categoría y porcentaje (respecto al total de futuros profesores)

Categorías	Contraste		Intervalo	
	N. profesores	%	N. profesores	%
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

Fuente: elaborado por el autor.

Seguidamente, clasificamos 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.

4.1 Errores en el planteamiento de las hipótesis

Los 55 errores que se citaron en esta categoría, exclusiva del contraste de hipótesis, se han clasificado tal y como se muestra a continuación:

Confunde hipótesis nula y alternativa. Este error, descrito por Vallecillos (1999) y Vera et al. (2011), conlleva realizar incorrectamente el resto del procedimiento, intercambiando las regiones críticas y de aceptación y, concluyendo con una decisión errónea:

Tomar la hipótesis a estudiar como hipótesis nula y no como alternativa, ya que la hipótesis nula no se acepta, si no que se rechaza o no se rechaza (Respuesta de LAP).

Plantea hipótesis que no complementarias. Este error, recogido en investigaciones realizadas con estudiantes (ESPINEL; RAMOS; RAMOS, 2007; VALLECILLOS, 1999) implica que las hipótesis del contraste no cubrirían en su totalidad los valores posibles del espacio paramétrico, por lo que el resultado podría no ser concluyente.

Poner la hipótesis nula y alternativa de manera que no sean complementarias (Respuesta de ILG).

Respuesta imprecisa. Algunos estudiantes dan respuestas imprecisas, como la

mostrada en el siguiente ejemplo en el que 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 3 se desglosan los 55 errores asociados al planteamiento de la hipótesis, la mayor parte de los cuales se relacionan con la confusión entre la hipótesis nula y alternativa.

Tabla 3 - Número de errores citados en el planteamiento de las hipótesis y porcentaje (respecto al total de estudiantes)

Categorías	N. Errores	Porcentaje
Confunde hipótesis nula y alternativa	35	50,0
Plantea hipótesis no complementarias	6	8,6
Respuesta imprecisa	14	20,0

Fuente: elaborado por el autor.

Es menos frecuente reconocer que las hipótesis podrían no ser complementarias y un quinto de los futuros profesores son imprecisos al citar errores sobre el planteamiento de las hipótesis. No se describen 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 (ESPINEL; RAMOS; RAMOS, 2007; HARRADINE et al., 2011; RAMOS; ESPINEL; RAMOS, 2009; VALLECILLOS, 1999; VERA et al., 2011).

4.2 Errores en la selección o planteamiento del contraste o del intervalo

En segundo lugar, clasificamos los errores descritos en referencia a la correcta selección de un contraste de hipótesis o un intervalo de confianza adecuado.

Error al identificar la distribución muestral. La selección de la distribución muestral adecuada es fundamental tanto en la realización de un contraste como en el cálculo de un intervalo de confianza, debido a que los cálculos posteriores, por ejemplo, el p -valor, o la amplitud del intervalo de confianza, se basan en ella. El error en la elección de la distribución muestral (por ejemplo, utilizar la distribución t en lugar de la normal o tomar unos parámetros erróneos para la distribución normal) es señalado por algunos estudiantes, como ACG, que indica que se puede confundir la distribución muestral, al expresar el intervalo en función de la distribución normal $N(0,1)$.

Una vez tipificado, el alumno puede olvidar deshacer el cambio y presentar el resultado final con la variable tipificada (Respuesta de ACG).

Error al identificar el tipo de contraste o de intervalo. Una elección incorrecta del tipo de contraste o intervalo (unilateral o bilateral) implica errores posteriores. Recordemos que, si

el contraste es bilateral, se establecen dos regiones con probabilidades $\alpha/2$ respectivamente, mientras que si este es unilateral, únicamente se cuenta con una región de rechazo con probabilidad α . El mismo problema se presenta entre intervalo de confianza unilateral o bilateral. Este error fue descrito por Espinel, Ramos y Ramos (2007).

No entender qué tipo de contraste deben hacer; si bilateral, unilateral por la izquierda o por la derecha (Respuesta de ALM).

De la información recogida en la Tabla 4 se observa que pocos participantes han señalado posibles errores en la elección del contraste o intervalo adecuado, y la mayoría de los que los citan en el contraste de hipótesis son imprecisos, lo que indica que los participantes en el estudio son poco conscientes de que sus alumnos podrían cometer estos errores.

Tabla 4 - Número de errores citados en el planteamiento y porcentaje (respecto al total de estudiantes)

Categorías	Contraste		Intervalo	
	N. Errores	%	N. Errores	%
Error al identificar la distribución muestral	8	11,4	5	7,1
Error al identificar el tipo de contraste (bilateral o unilateral)	8	11,4	4	5,7
Respuesta imprecisa	12	17,1	2	2,9

Fuente: elaborado por el autor.

4.3 Errores conceptuales

Son muchos los conceptos que intervienen tanto en el contraste de hipótesis como en el intervalo de confianza. Respecto a los conceptos comunes a ambos procedimientos, los participantes en el estudio citaron los siguientes errores:

Confunde estadístico y parámetro. En esta categoría se han considerado aquellas respuestas en las que los estudiantes confunden los parámetros poblacionales (resúmenes desconocidos y constantes de una población) con los estadísticos (valores conocidos y que varían de una muestra a otra) (VALLECILLOS, 1999; VERA et al., 2011). Este error está relacionado con la falta de discriminación de 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). En el siguiente ejemplo, un participante en el estudio cita este error:

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).

Comprensión incorrecta del Teorema Central del Límite. Este teorema asegura que la

suma de variables aleatorias independientes, con varianza finita, tiende a distribuirse normalmente cuando aumenta el tamaño muestral. Puesto que la media muestral se calcula como suma de variables aleatorias, permite al aplicarlo, identificar la distribución muestral, si el tamaño de la muestra es adecuado y el muestreo es aleatorio. Además, se deben asignar unos valores correctos a la media y desviación típica de la distribución normal resultante, lo que es resaltado por algunos futuros profesores. Aunque son varios los estudiantes que citan este error, las respuestas son algo ambiguas ya que únicamente se indica la no correcta aplicación del mismo o el desconocimiento de su uso.

No saber cómo utilizar el Teorema Central del Límite (Respuesta de MCM)

Interpreta incorrectamente el nivel de significación o el nivel de confianza. Algunos futuros profesores sugieren errores de comprensión del nivel de significación, aunque no suelen ser precisos para describir 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 error al rechazar la hipótesis nula, u otro de los errores asociados a dicho concepto (FALK, 1986, VALLECILLOS, 1999). Por tanto, no queda claro si realmente los participantes conocen los errores típicos asociados a este concepto.

Igual ocurre con los estudiantes que indican errores en la interpretación del nivel de confianza. Mostramos a continuación la respuesta, muy imprecisa, dada por MHB, que sugiere un error no citado en las investigaciones previas, consistente en confundir el nivel de significación con su complementario (nivel de confianza).

Confundir el nivel de significación con el nivel de confianza (Respuesta de MHB)

Confunde otros conceptos: se considera aquellos casos en los que se sugiere la confusión entre conceptos propios de estadística descriptiva (como confundir desviación típica y varianza) o relacionados con la estimación puntual; por ejemplo, confundir el estimador de la proporción con su complementario.

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).

Adicionalmente, se citaron los siguientes errores específicos del contraste de hipótesis.

Interpreta incorrectamente el p-valor. El *p*-valor se define como la probabilidad de obtener el valor observado del estadístico u otro más extremo, supuesta cierta la hipótesis nula. Dicho concepto es interpretado por algunos estudiantes como la probabilidad de que la hipótesis sea cierta si se obtuvo el valor observado del estadístico, como la probabilidad de

que el valor obtenido del estadístico se deba al azar o como la probabilidad de que la hipótesis nula sea cierta (CASTRO-SOTOS et al., 2007). Ninguno de estos errores fue citado por los futuros profesores, pues sólo MSM sugiere la confusión entre el p -valor y el nivel de significación. Esta confusión es importante, pues mientras que 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).

Confunde los errores Tipo I y Tipo II. Una vez tomada la decisión sobre la hipótesis nula, existe la posibilidad de cometer errores dependiendo de la decisión tomada. Mientras que el error Tipo I consiste en rechazar la hipótesis nula siendo cierta, el Error Tipo II reside en no rechazar la hipótesis nula cuando es falsa. La importancia de diferenciarlos se debe a que la probabilidad asociada al primer error (nivel de significación) queda determinada antes del contraste, y la probabilidad asociada al segundo es variable y dependiente de diversos factores. Un ejemplo de cita de este error se reproduce a continuación:

Confundir el error Tipo I con el error Tipo I (Repuesta de JGM).

La Tabla 5 recoge los errores conceptuales citados por los futuros profesores.

Tabla 5 – Número de errores conceptuales citados y porcentaje (respecto al total de futuros profesores)

Categorías	Contraste		Intervalo	
	N. Errores	%	N. Errores	%
Confunde estadístico y parámetro	30	42,9	20	28,6
Comprensión incorrecta del Teorema Central del Límite	4	5,7		
Interpreta incorrectamente nivel de significación o de confianza	6	8,6	4	5,7
Confunde otros conceptos	1	1,4	4	5,7
Interpreta incorrectamente el p -valor	2	2,9		
Confunde los errores Tipo I y Tipo II	10	14,3		

Fuente: elaborado por el autor.

La Tabla 5 muestra que, en ambas tareas, fue frecuente citar el error asociado a la confusión entre parámetro y estadístico, en especial, en el contraste de hipótesis, donde los participantes lo ven más previsible que en la construcción de un intervalo de confianza. El segundo error conceptual más citado es el relacionado con los tipos de errores en un contraste de hipótesis; de hecho, este es un error no evitable; es consustancial al contraste de hipótesis, mientras que el resto se debe a un aprendizaje deficiente del procedimiento y pueden corregirse.

El resto de errores conceptuales son citados con poca frecuencia. Asimismo, observamos que los errores asociados a la aplicación del Teorema Central del Límite no son

considerados en la construcción del intervalo de confianza. Por último, resaltar que el número de citas de errores conceptuales asociados al contraste de hipótesis casi 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.

Destacamos que apenas se ha aludido a los errores de interpretación del nivel de significación y del p -valor, lo que nos hace sospechar que los participantes pudieran no ser conscientes de los errores relacionados con la interpretación de estos conceptos, que, de acuerdo a Badenes-Ribera et al. (2015), Batanero (2000) y Caperos y Pardo (2013), están muy extendidos.

4.4. 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 citados en el presente trabajo se analizan a continuación:

Error en la identificación de los datos. El primer paso del procedimiento, citado por algunos futuros profesores, 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.

Interpretar los datos que nos da el problema. (Respuesta de LCM).

Determinación del estadístico de contraste. La realización de un contraste, o la construcción de un intervalo, requiere la elección de un estadístico muestral en relación al parámetro que se desea estimar o contrastar. Por ejemplo, el intervalo de confianza de la media poblacional se fundamenta en el uso de la media muestral, estimador insesgado de la misma. Un error señalado por algunos participantes es la elección inadecuada del estadístico.

Confundir el estadístico que hay que calcular para comprobar con el valor teórico. (Respuesta de JRC).

Error de tipificación o al operar con desigualdades. En el cálculo asociado al intervalo de confianza o al contraste de hipótesis, 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 describen Espinel et al. (2007) y lo recoge la respuesta JLS:

También pueden equivocarse en un paso clave (si fuera necesario): tipificar la variable. (Respuesta JLS)

Error en el cálculo de probabilidades. En ocasiones se citan los errores de cálculo de probabilidades, como, por ejemplo, al calcular probabilidades de que la variable en estudio esté comprendida entre dos valores. Este hecho 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:

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).

Lectura incorrecta de las tablas estadísticas. 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:

No saber buscar en las tablas para cada distribución. (Respuesta de MVV).

Error en el cálculo del p-valor. El cálculo del p-valor puede ser erróneo cuando se calcula como probabilidad simple, en lugar de condicionada, error descrito por Falk (1986), o bien si, en lugar de tomar la desigualdad apropiada para el tipo de contraste, se toma la contraria. Los estudiantes que indican este error procedimental suelen ser muy imprecisos, como el siguiente:

Pueden errar en el cálculo e incluso la interpretación del p-valor correspondiente a la observación. (Respuesta de JLS).

Determinación del punto crítico. Mientras que el cálculo del p-valor parte de un valor observado del estadístico para calcular su probabilidad, la determinación del punto crítico se realiza mediante la operación inversa, pues se toma un valor de probabilidad (nivel de significación) para determinar el valor del estadístico que corresponde a dicha probabilidad, lo que es citado en forma confusa por ciertos futuros profesores.

Podrían cometer errores al calcular el valor crítico $Z_{1-\alpha}$. (Respuesta de AGB).

Determinación incorrecta de las regiones de rechazo/no rechazo. La región de rechazo está formada por un conjunto de valores posibles del estadístico de contraste alejados del valor supuesto en 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), y citado en el siguiente ejemplo:

En los contrastes unilaterales en lugar de calcular $Z_{1-\alpha}$, calcula $Z_{\alpha/2}$, probablemente por la inercia que tiene para calcular intervalos de confianza donde calcula $Z_{\alpha/2}$. En los contrastes unilaterales, suele equivocarse a la hora de tomar la región de aceptación, al no tener en

cuenta el sentido de la desigualdad de la hipótesis nula. (Respuesta de MVC).

Errores al establecer los límites del intervalo. La obtención del extremo inferior y superior del intervalo de confianza implica restar y sumar al estadístico el error de estimación, respectivamente. Consecuentemente, un intervalo puede quedar incorrectamente determinado si alguno de estos valores no es adecuado o si el extremo inferior se obtiene como resultado de la suma y el superior como resultado de la resta. A modo de ejemplo se tiene la siguiente respuesta.

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).

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 en el uso de la notación.

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).

Tabla 6 – Número de errores procedimentales citados y porcentaje (respecto al total de futuros profesores)

Categorías	Contraste		Intervalo	
	N. Errores	%	N. Errores	%
Error en la identificación o interpretación de datos	7	10,0	16	22,9
Determinación del estadístico de contraste	10	14,3	3	4,3
Error de tipificación o al operar con desigualdades	18	25,7	38	54,3
Error en el cálculo de probabilidades	5	7,1	2	2,9
Lectura incorrecta de las tablas estadísticas	23	32,9	21	30,0
Error en el cálculo del p -valor	2	2,9		
Determinación del punto crítico	6	8,6		
Determinación incorrecta de las regiones	9	12,9		
Error al establecer los límites del intervalo			10	14,3
Errores de cálculo o en la terminología	10	14,3	5	7,1

Fuente: elaborado por el autor.

Los errores procedimentales citados en ambas tareas (véase Tabla 6) son bastante variados y numerosos, aunque la mayor parte de las veces hay poca precisión al describirlos. Los errores procedimentales citados con mayor frecuencia son los relacionados con la lectura de las tablas de la distribución normal y tipificación, identificados por Espinel y Ramos (2007), y determinación de las regiones.

4.5. Errores de interpretación

La última etapa en la realización de un contraste de hipótesis o en el cálculo de un intervalo de confianza es la interpretación de los resultados de los cálculos matemáticos, en el

caso del contraste de hipótesis para tomar una decisión sobre la misma y en los dos procedimientos para contextualizar en el contexto del problema. Respecto al contraste de hipótesis se citan los siguientes errores de interpretación:

No ser capaz de tomar una decisión sobre la hipótesis nula. Incluso realizando correctamente el procedimiento asociado a un contraste de hipótesis, se puede no llegar a tomar la decisión adecuada, si no comprende a fondo el proceso por el cual se toma la decisión. Este es un error poco citado en la literatura.

Error a la hora de rechazar la hipótesis nula o no rechazar la hipótesis nula. (Respuesta de CBF).

Aceptar la hipótesis nula. El contraste de hipótesis es un procedimiento diseñado para proporcionar evidencia empírica que permite contradecir la hipótesis nula. Por ello, si la hipótesis nula no llega a rechazarse, sólo se puede indicar que los datos de la muestra no ofrecen suficientes evidencias sobre su falsedad (Batanero, 2000). Por el contrario, si los datos contradicen la hipótesis, esta se rechaza. Algunos futuros profesores indican la posible confusión que al respecto podrían tener los estudiantes:

Igualmente pueden producirse errores en la formulación de lenguaje matemático en cuanto al uso de las palabras "rechazo" y "aceptación" de hipótesis, ya que debe saber que un contraste no establece la veracidad de una hipótesis. (Respuesta de BRR).

Cometer un error tipo I o un error tipo II. Debido a la aleatoriedad presente en el muestreo, existen dos tipos de errores; mantener la hipótesis nula siendo falsa (Error Tipo II) o bien rechazarla siendo verdadera (Error Tipo I). Los futuros profesores han aludido a este tipo de error en algunas de sus respuestas.

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).

No se contextualiza 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 por 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:

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.

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).

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).

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).

Considerar el parámetro dentro del intervalo. Otro error frecuente es considerar que el valor del parámetro está dentro de cualquier intervalo de confianza, aunque en realidad sólo un porcentaje de intervalos (dado por el coeficiente de confianza) lo contendrá (CUMMING et al., 2004). En la siguiente respuesta se refleja este error.

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 de ACG).

Aceptar el intervalo sin comprobar que contiene a la media verdadera. Por otro lado, algunos participantes en el estudio consideran erróneamente que el intervalo de confianza contiene con seguridad al parámetro poblacional y suponen que un error en el estudiante es no interpretar el intervalo en esta forma:

Aceptar sin comprobar que los intervalos de confianza contienen la media verdadera en el 95% de los casos (Respuesta de MMB).

Interpretación incorrecta de factores que afectan al resultado. Por último, y común a los dos métodos puede darse una interpretación incorrecta de 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).

La Tabla 7 resume los errores citados sobre interpretación donde se observa que un alto porcentaje de futuros profesores es consciente de que los extremos de un intervalo de confianza son aleatorios. También se percibe la importancia que los futuros profesores le otorgan a la contextualización de los resultados y a la interpretación del intervalo de confianza. Con mucha menor frecuencia se citan el resto de errores, como por ejemplo, no ser capaz de tomar una decisión sobre el rechazo de la hipótesis o aceptar una hipótesis nula.

Tabla 7 – Número de errores de interpretación y porcentaje (respecto al total de estudiantes)

Categorías	Contraste		Intervalo	
	N. Profesores	%	N. Profesores	%
No ser capaz de tomar una decisión sobre la hipótesis nula	16	22,9		
Aceptar la hipótesis nula	8	11,4		
Cometer un error Tipo I o Tipo II	14	20,0		
No se contextualiza el resultado	31	44,3		
Interpretación del nivel de confianza o significación			12	17,1
Considerar fijos los extremos del intervalo			31	44,3
Considerar el parámetro dentro del intervalo			6	8,6
Aceptar el intervalo sin comprobar que contiene a la media verdadera			5	7,1
Interpretación incorrecta de factores que afectan el resultado	1	1,4	7	10,0

Fuente: elaborado por el autor.

5 Conclusiones

Los participantes del estudio fueron capaces de identificar un número apreciable de errores, tanto en el desarrollo del contraste de hipótesis como en la construcción de intervalos de confianza por parte de sus futuros estudiantes. Ello muestra también un conocimiento matemático razonable del tema, pues son capaces de predecir qué pasos serán erróneos. Sin embargo, 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 Badenes-Ribera et al. (2015), 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 valor p . 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 Inzusa y Jiménez (2013) y de sus propios errores conceptuales en el desarrollo del contraste de hipótesis (LÓPEZ-MARTÍN; BATANERO; GEA, 2018).

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, Ramos y Ramos (2007) o Vera et al. (2011).

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, Espinel y Ramos (2009) y la confusión entre estadístico y parámetro (HARRADINE et al., 2011).

Sin embargo, muchos futuros profesores describen los errores con poca precisión, lo que nos indica que su conocimiento del tema es todavía escaso; por ejemplo, cuando se alude a hipótesis mal planteadas, a la interpretación incorrecta del intervalo de confianza o no saber plantear bien el contraste, sin indicar concretamente en qué consiste el error.

Una implicación didáctica es que, aunque se observa un cierto grado de desarrollo de la faceta cognitiva del conocimiento didáctico-matemático (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. Por otro lado, los futuros profesores necesitan adquirir conocimiento sobre niveles de desarrollo del razonamiento inferencial de sus estudiantes, como los descritos por

Inzunza y Jimenez (2013) y sobre estrategias alternativas de resolver correctamente los problemas.

Además, sería necesario desarrollar en ellos, también, el resto de las facetas de su conocimiento didáctico-matemático sobre este tema. En particular, sería importante acrecentar su faceta instruccional para proporcionarle modelos y situaciones didácticas que ayuden a mejorar el conocimiento de los estudiantes sobre el contraste de hipótesis con el fin de 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, Díaz-Batanero y López-Martín (2017) o Wild et al. (2011) no hay una única aproximación a la inferencia estadística y es posible hoy día, gracias a la simulación, comenzar con aproximaciones informales. En estas aproximaciones (véase, por ejemplo, NOLL; GEBRESENBET; GLOVER, 2016; o RUBIN; HAMMERMAN; KONOLD, 2006), para resolver los problemas, se sustituye la distribución muestral teórica normal, por una distribución muestral empírica, obtenida mediante simulación. Los estudiantes, entonces, superan las posibles dificultades asociadas a la tipificación y uso de las tablas de la distribución normal, y se pueden concentrar en el aprendizaje de conceptos como hipótesis nula y alternativa, región crítica y de aceptación, nivel de significación y de confianza o intervalo de confianza.

Todo ello teniendo en cuenta que una aproximación informal, por sí sola, no asegura que los estudiantes no cometan errores conceptuales (SILVESTRE; SÁNCHEZ, 2016), por lo que es necesario discutir con los estudiantes (o futuros profesores en caso de que se trate de su formación) los conceptos subyacentes a lo largo de los procedimientos y reforzar, igualmente, las actividades de interpretación y contextualización de los resultados.

Agradecimientos

Proyecto EDU2016-74848-P (AEI, FEDER) y Grupo FQM126 (Junta de Andalucía).

Referencias

BADENES-RIBERA, L. et al. Interpretation of the p -value: A national survey study in academic psychologist from Spain. **Psicothema**, Universidad de Oviedo, España, v. 27, p. 290 -295, 2015.

BATANERO, C. Controversies around the role of statistical tests in experimental research.

Mathematical Thinking and Learning, Queensland University of Technology, Australia. v. 2, n. 1-2, p. 75-98, 2000.

BATANERO, C.; DÍAZ, C; LÓPEZ-MARTÍN, M. M. Significados del contraste de hipótesis, configuraciones epistémicas asociadas y algunos conflictos semióticos. In: Contreras, J. M.; Arteaga, P.; Cañadas, G. R.; Gea, M. M.; Giacomone, B.; López-Martín, M. M. (Eds.). **Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos**. Granada, 2017. Online: <<http://enfoqueontosemiotico.ugr.es/civeos/batanero.pdf>>. Acceso: 5 jun. 2019.

BEHAR, R. **Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística**. 2001. 210 f. Tesis (Doctoral en Estadística) - Universidad Politécnica de Cataluña, España, 2001.

BIEHLER, R.; FRISCHEMEIER, D; PODWORN, S. Preservice teachers' reasoning about uncertainty in the context of randomization tests. In: A. S. ZIEFFLER, A. S.; FRY, E. (Ed.). **Reasoning about uncertainty: learning and teaching informal inferential reasoning**. Catalyst Press, 2015. p. 129 - 162.

CAPEROS, J. M.; PARDO, A. Consistency errors in p -values reported in Spanish psychology journals. **Psicothema**, Universidad de Oviedo, España, v. 25, p. 408 - 414, 2013.

CASTRO-SOTOS, A. E. et al. Student's misconceptions of statistical inference: A review of the empirical evidence from research on statistical education. **Educational Research Review**, University of Regensburg, Alemania, v. 2, n. 2, p. 98 - 113, 2007.

CUMMING, G.; WILLIAMS, J.; FIDLER, F. Replication and researchers' understanding of confidence intervals and standard error bars. **Understanding statistics**, Taylor & Francis, Abingdon, Inglaterra, v. 3, n. 4, p. 299 - 311, 2004.

DAWSON, A. J.; JAWORSKI, B.; WOOD, T. (Ed.). **Mathematics teacher education: Critical international perspectives**. Dordrecht, The Netherlands: Routledge, 2013.

ESPINEL, M. C.; RAMOS, R. M.; RAMOS, C. E. Algunas alternativas para la mejora de la enseñanza de la inferencia estadística en Secundaria. **Números**, La Laguna, Tenerife, España, v. 67, p. 15 - 23, 2007.

FALK, R. Conditional probabilities: insights and difficulties. In: Davidson, R.; Swift, J. (Ed.). **INTERNATIONAL CONFERENCE ON TEACHING STATISTICS, 2.**, 1986. Victoria. **Proceedings...** Victoria, Canada: International Statistical Institute, 1986. p. 292 - 297.

FIDLER, F.; CUMMING, G. Teaching confidence intervals: Problems and potential solutions. In: **INTERNATIONAL STATISTICAL INSTITUTE, 55.**, 2005. Voorburg. **Proceedings...** Voorburg: International Statistical Institute, 2005. p. 1-5. Online: <iase-web.org/documents/papers/isi55/Fidler-Cumming.pdf>. Access: 8 Febrero 2018.

GODINO, J. D. Categorías de análisis de los conocimientos del profesor de matemáticas. **UNIÓN**, Buenos Aires, Argentina, v. 20, p. 13 - 31, 2009.

GODINO, J. D.; BATANERO, C.; FONT, V.; GIACOMONE, B. Articulando conocimientos y competencias del profesor de matemáticas: el modelo CCDM. In: Macías, J. A.; Jiménez, A.; González, J. L.; Sánchez, M. T.; Hernández, P.; Fernández, C.; Ruiz, F. J.; Fernández, T.; Berciano, A. (Ed.). **Investigación en Educación Matemática XX**. Málaga: SEIEM, 2016, p. 285-294.

HALLER, H. ; KRAUSS, S. Misinterpretations of significance: A problem students share with their teachers? **Methods of Psychological Research**, INSERTAR LUGAR DE LA PUBLICACIÓN, v. 7, n. 1, p. 1 - 20, 2002.

HARRADINE, A.; BATANERO, C.; ROSSMAN, A. Students' and teachers' knowledge of sampling and inference. In: Batanero, C.; Burrill, G.; Reading, C. (Eds.). **Teaching statistics in school mathematics- Challenges for teaching and teacher education. A Joint ICM/IASE Study**. New York: Springer, 2011.

HOEKSTRA, R.; MOREY, R. D.; ROUDER, J. N.; WAGENMAKERS, E. J. Robust misinterpretation of confidence intervals. **Psychonomic Bulletin & Review**, Irvine, California, v. 21, n. 5, p. 1157 - 1164, 2014.

INZUNSA, S.; JIMÉNEZ, J. V. Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. **Revista Latinoamericana de Investigación en Matemática Educativa**, México D.F., v. 16, n. 2, p. 179 - 211, 2013.

KRAINER, K. Teachers as stakeholders in mathematics education research. **The Mathematics Enthusiast**, Ottawa, Canada, v. 11, n. 1, p. 49 - 60, 2014.

LIU, Y.; THOMPSON, P. W. Mathematics teachers' understandings of proto-hypothesis testing. **Pedagogies**, Taylor & Francis, Abingdon, Inglaterra, v. 4, n. 2, p. 126 - 138, 2009.

LLINARES, S. Mathematics teacher's knowledge, knowledge-based reasoning, and contexts. **Journal of Mathematics Teacher Education**, Atenas, Grecia, v. 21, n. 1, p. 1 - 3, 2018.

LÓPEZ-MARTÍN, M. M.; BATANERO, C.; DÍAZ-BATANERO, C.; GEA, M. La inferencia estadística en las Pruebas de Acceso a la Universidad en Andalucía, **Revista Paranaense de Educação Matemática**, Paraná, Argentina, v. 5, n. 8, p. 33-59; 2016.

LÓPEZ-MARTÍN, M. M.; BATANERO, C.; GEA, M. La faceta cognitiva del conocimiento de futuros profesores sobre el contraste de hipótesis. **Investigación en Educación Matemática XXII**. Gijón, p. 300-309, 2018.

MAKAR, K.; RUBIN, A. Learning about statistical inference. In: Ben-Zvi, D.; Garfield, J. B.; Makar, K. (Ed.). **International Handbook of Research in Statistics Education**, Nueva York: Springer, 2018. p. 261 - 294. https://doi.org/10.1007/978-3-319-66195-7_8

MCLEAN, A. Statistacy: Vocabulary and hypothesis testing. In: Phillips, B. (Ed.). **INTERNATIONAL CONFERENCE ON TEACHING STATISTICS, 6.**, 2002. Cape Town. **Proceedings...** Cape Town: International Statistical Institute, 2002. Online <https://iase-web.org/documents/papers/icots6/3m2_mcle.pdf>Access: 5 Junio 2019.

MEC, MINISTERIO DE EDUCACIÓN Y CIENCIA. **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, 2007.

MECD, MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE. **Real Decreto 1105/2014, de 26 de diciembre por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato**. Madrid: Autor, 2015.

NEUENDORF, K. A. **The content analysis guidebook**. London: Sage, 2016.

NOLL, J.; GEBRESENBET, M.; GLOVER, E. D. A modeling and simulation approach to informal inference: successes and challenges. En D. Ben-Zvi (Ed.). **The teaching and learning of statistics**. New York: Springer, 2016. p. 139 - 150.

PINO-FAN, L.; GODINO, J. D. Perspectiva ampliada del conocimiento didáctico-matemático del profesor. **Paradigma**, Caracas, Venezuela, v. 36, n. 1, p. 87 - 109, 2015.



RAMOS, C. E.; ESPINEL, M. C.; RAMOS, R. M. Identificación de los errores en los contrastes de hipótesis de los alumnos de Bachillerato. **SUMA**, Zaragoza, España, v. 61, p. 35 - 44, 2009.

RUBIN, A.; HAMMERMAN, J. K. L.; KONOLD, C. Exploring informal inference with interactive visualization software. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education. Online: <www.stat.auckland.ac.nz/~iase/publications>. Acceso 5 junio 2019. 2006

SILVESTRE, E.; SÁNCHEZ, E. Patrones en el desarrollo del razonamiento inferencial informal: introducción a las pruebas de significancia en el bachillerato. In: Macías, J. A.; Jiménez, A.; González, J. L.; Sánchez, M. T.; Hernández, P.; Fernández, C.; Ruiz, F. J.; Fernández, T.; Berciano, A. (Ed.). **Investigación en Educación Matemática XX**, Málaga: SEIEM, 2016. p. 509-518.

VALLECILLOS; A. Some empirical evidence on learning difficulties about testing hypotheses. **Bulletin of the International Statistical Institute**, Londres, v. 58, p. 201-204, 1999.

VERA, O.; DÍAZ-BATANERO, C.; BATANERO, C. Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. **Unión**, v. 27, p. 41-61, 2011.

WILD, C. J. et al. Towards more accessible conceptions of statistical inference (with discussion). **Journal of the Royal Statistical Society (A)**, Londres, v. 174, n. 2, p. 247-295, 2011.

Submetido em 24 de Maio de 2018.
Aprovado em 17 de Outubro de 2018.