Publicaciones. Facultad de Educación y Humanidades del Campus de Melilla
Depósito legal: GR-94-2001 · eISSN: 2530-9269 · pISSN: 1577-4147
Edita: Facultad de Ciencias de la Educación y del Deporte de Melilla (Universidad de Granada)

Conocimiento del contraste de hipótesis por futuros profesores de Educación Secundaria y Bachillerato

Knowledge of Statistical Test by Prospective Secondary and High School Teachers

Carmen Batanero Bernabeu
Universidad de Granada
batanero@ugr.es
https://orcid.org/0000-0002-4189-7139

Mª del Mar López-Martín
Universidad de Granada
mariadelmarlopez@ugr.es
https://orcid.org/0000-0001-8677-9606

Mª Magdalena Gea Serrano
Universidad de Granada
mmgea@ugr.es
https://orcid.org/0000-0002-5229-0121

Pedro Arteaga Cezón
Universidad de Granada
parteaga@ugr.es
https://orcid.org/0000-0002-8347-7669

Fechas · Dates

Recibido: 2017-03-04
Aceptado: 2018-05-21
Publicado: 2018-12-27

Cómo citar este trabajo · How to Cite this Paper

Batanero, C., López-Martín, M. M., Gea, M. M., & 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

Resumen

El objetivo de este trabajo fue evaluar los conocimientos de un grupo de futuros profesores de Educación Secundaria y Bachillerato sobre el contraste de hipótesis. Para ello se analizaron las respuestas de un grupo de 73 futuros profesores del Máster de Educación Secundaria de la Universidad de Granada, a un problema abierto, similar a los propuestos en cursos anteriores en las pruebas de acceso a la universidad. Los resultados indican que, aunque la mayoría de los participantes en el estudio planteó correctamente las hipótesis y eligió un contraste consistente con las mismas, sólo el 15% realizó todo el procedimiento sin errores, y únicamente el 38,4% tomó la decisión correcta y contextualizó, además, los resultados. Se detectaron errores similares a los descritos en investigaciones previas con estudiantes, lo que plantea la necesidad de mejorar la formación de los futuros profesores sobre este contenido.

Palabras clave: Contraste de hipótesis; conocimiento del contenido; futuros profesores.

Abstract

The aim of this research was to assess prospective secondary and high school teachers' knowledge of hypothesis tests. To achieve this aim the responses given by 73 students in the Master of Secondary Education at the University of Granada to an open problem similar to those included in the previous years at the entrance to university tests are analysed. The results suggest that, although the majority of participants set correct hypotheses and selected a test consistent with the same, only 15% of them correctly completed the procedure, and just a 38,4% made the correct decision and contextualized, moreover, the results. Errors similar to those described in previous research with students were found, which suggests the need of improving the prospective teachers' preparation on this content.

Key words: Statistical tests; content knowledge; prospective teachers.

Introducción

La inferencia estadística tiene un papel destacado en las investigaciones en ciencias humanas y, en consecuencia, las orientaciones curriculares recientes de Bachillerato de Humanidades y Ciencias Sociales (MECD, 2015) proponen contenidos sobre muestreo, teorema central del límite e intervalos de confianza. Además, en el currículo anterior (MEC, 2007), que es el que tenemos en cuenta en este trabajo, se consideró, junto a los contenidos mencionados anteriormente, el contraste de hipótesis sobre la media de una distribución normal y la proporción de una distribución binomial. Igualmente, las pruebas de acceso a la universidad (Pruebas PAU) de Matemáticas Aplicadas a las Ciencias Sociales en Andalucía propusieron en los últimos 15 años un problema (entre cuatro) sobre inferencia estadística, incluyendo problemas de contraste de hipótesis (López-Martín, Batanero, Díaz-Batanero y Gea, 2016).

McLean (2002) resalta la importancia de unos conocimientos básicos de contraste de hipótesis, como parte de la cultura estadística de un ciudadano bien formado. Sin embargo, este tema no es sencillo, y ejemplo de ello son las críticas que se han realizado en torno al uso incorrecto del contraste de hipótesis (por ejemplo, en Gigerenzer, 1993, Harlow, Mulaik y Steiger, 1997; Morrison y Henkel, 2006; Nickerson, 2000) o los errores descritos en la comprensión de la inferencia (Castro Sotos, Vanhoof, Van den

Nororgate y Onghena, 2007; Haller y Krauss, 2002; Vera, Díaz y Batanero, 2011). Ante esta situación, Harradine, Batanero y Rossman (2011) añaden que, mientras que la realización de los cálculos asociados con los contraste de hipótesis es hoy día muy sencilla gracias al software estadístico, los estudiantes continúan teniendo dificultades en la comprensión de los conceptos asociados.

La mejora de la comprensión y la aplicación de la inferencia implican la preparación adecuada de los profesores responsables de su enseñanza. Sin embargo, aunque algunas investigaciones han estado centradas en la comprensión de conceptos de inferencia por futuros profesores, el contraste de hipótesis apenas ha sido tenido en cuenta.

El objetivo de este trabajo es realizar un estudio exploratorio de evaluación del conocimiento del contraste de hipótesis de los futuros profesores que se preparan como profesores de matemáticas de Secundaria y Bachillerato. Consideramos únicamente el conocimiento matemático del tema. Con ello completamos las escasas investigaciones previas, y proporcionamos información de ayuda al formador de profesores. A continuación presentamos los fundamentos del trabajo, el método y los resultados, concluyendo con unas sugerencias para la formación de profesores.

Investigación previa

La realización de un contraste de hipótesis está influenciada por la comprensión de una gran diversidad de conceptos, lo que explica la variedad de posibles errores asociados (Harradine et al., 2011; Inzunsa y Jiménez, 2013; Vallecillos, 1994; 1999).

Entre los conceptos que suelen tener una interpretación incorrecta destaca el nivel de significación, denotado habitualmente como α y definido como la probabilidad de rechazar la hipótesis nula (H_0), supuesta cierta, es decir, $\alpha = P(\text{rechazar } H_0 | H_0 \textit{ es cierta})$. La interpretación errónea más común del nivel de significación consiste en cambiar los dos términos de la probabilidad condicionada en la expresión anterior, es decir, interpretar α como la probabilidad de que la hipótesis nula sea cierta, una vez que la decisión de rechazarla se ha tomado, esto es, interpretar $\alpha = P(H_0 \text{ es cierta} \mid \text{se ha rechazado } H_0)$. Este error se encuentra ya recogido en las investigaciones llevadas a cabo por Birnbaum (1982) y Falk (1986). El primer autor pidió a sus estudiantes que definiesen el nivel de significación, obteniendo como respuesta más frecuente la siguiente: *Un nivel de significación del 5% significa que, en promedio, 5 de cada 100 veces que rechacemos la hipótesis nula estaremos equivocados*. De manera similar, Falk (1986) señaló que la mayoría de sus estudiantes creían que α era la probabilidad de equivocarse al rechazar la hipótesis nula. Posteriormente, Vallecillos (1994; 1999) también encontró este error en una investigación con más de 400 estudiantes de la Universidad de Granada de diversas especialidades (matemáticas, medicina, ingeniería o psicología). Resultados similares fueron encontrados por Haller y Krauss (2002) al interpretar el *p*-valor o probabilidad de encontrar un valor del estadístico muestral igual o más extremo que el observado, suponiendo la hipótesis nula cierta; este error también se encontró en profesores responsables de la enseñanza de métodos de investigación y en publicaciones en el campo de la psicología (Badenes-Ribera, Frías-Navarro, Monterde-i-Bort y Pascual-Soler, 2015; Caperos y Pardo, 2013).

Otros errores destacados son los referidos a la comprensión del papel jugado por las hipótesis nula y alternativa en un contraste y la forma de plantearlas (Vallecillos, 1994). A la hora de llevar a cabo un contraste de hipótesis, la hipótesis nula, conocida también como hipótesis de no efecto, se plantea para ser rechazada, mientras que la alternativa sería la negación de la anterior (Batanero, 2000). En la realización del contraste, la hipótesis nula se supone cierta y la distribución muestral del estadístico se determina considerando este supuesto. Estas propiedades no las comprenden los estudiantes, como se observó en el trabajo de Vallecillos (1994), quien encontró que un 13% aproximadamente de 436 estudiantes confundieron la hipótesis nula con la alternativa al plantear las hipótesis en problemas de contraste de hipótesis sencillos. Además, observó que un 20% de los estudiantes plantearon sus hipótesis usando los estadísticos muestrales, por lo que no discriminaron si la hipótesis del contraste se refería al parámetro de la población o bien al estadístico muestral[1]. La confusión entre muestra y población y la dificultad para discernir entre parámetro, estimador y estimación, también han sido descritas por Espinel, Ramos y Ramos (2007), Harradine et al. (2011) y Ramos, Espinel y Ramos (2009). Por su parte Liu y Thompson (2009) realizaron entrevistas a ocho futuros profesores tratando de comprender sus dificultades con el contraste de hipótesis. Los sujetos mostraron una concepción determinista del contraste y falta de comprensión de la lógica del mismo. Los autores participantes en su estudio sólo consideraron la muestra dada en los ejercicios que resolvieron, sin ser capaces de imaginar las infinitas muestras potenciales que pueden tomarse de una población; por ello no consideraron la media de la muestra como una variable aleatoria.

Vera et al. (2011) analizan la formulación de hipótesis estadísticas en un contraste sobre la media de una población por 93 estudiantes de psicología, después de estudiar el tema, con la finalidad de identificar errores de comprensión de los conceptos que intervienen en el planteamiento de las hipótesis estadísticas. Entre otros errores se encontraron estudiantes que confunden el contraste unilateral y bilateral (error también descrito por Espinel et al., 2007), definen hipótesis que no cubren el espacio paramétrico y usan el estadístico muestral en vez del parámetro poblacional para definir las hipótesis.

Inzunsa y Jiménez (2013) realizan un estudio con 11 estudiantes voluntarios del grado de matemáticas en una universidad mexicana, que habían estudiado previamente el contraste de hipótesis, a los que pasa un cuestionario con algunos ítems abiertos y otros de opción múltiple. Utiliza las respuestas para situar a los 11 participantes en diferentes niveles de razonamiento estadístico según la taxonomía SOLO. Sus resultados indican que la mayoría de sus participantes se encontraban en un bajo nivel de razonamiento respecto al *p*-valor, nivel de significación e interpretación de los resultados de un contraste de hipótesis. Concluyen además, que los estudiantes poseen información aislada de los conceptos que intervienen en el contraste de hipótesis y no los ponen en relación. Igualmente algunos mostraron la concepción descrita por Vallecillos (1994), de que un contraste de hipótesis es sinónimo de una prueba matemática que establece la verdad de la hipótesis. Otros errores descritos por Espinel et al. (2007) cuando los estudiantes abordan el contraste de hipótesis son la tipificación errónea, el uso incorrecto de la tabla de probabilidades de la Normal y la interpretación errónea de las conclusiones en los contrastes.

Motivados por analizar estos errores, en este trabajo tratamos de confirmar si aparecen los mismos o diferentes en futuros profesores de Matemáticas de Educación Se-

1. Recordemos que mientras que el parámetro de la población (por ejemplo la media de una distribución normal) es un valor constante y desconocido, el correspondiente estadístico (en el ejemplo, la media de una muestra tomada de dicha población) es conocido (pues obtenida la muestra, se puede calcular) pero variable (ya que diferentes muestras tomadas de la misma población pueden dar lugar a valores ligeramente distintos de las medias muestrales).

cundaria y Bachillerato españoles, puesto que no hemos encontrado investigaciones previas sobre esta temática. Utilizando un problema abierto, se tendrán en cuenta el planteamiento de las hipótesis, selección del tipo de contraste, los cálculos realizados y la interpretación de los resultados obtenidos.

Metodología

El estudio se desarrolló con 73 estudiantes que se preparaban para ser futuros profesores de Secundaria y Bachillerato, dentro del Máster de Formación del Profesorado de Educación Secundaria y Bachillerato, especialidad de Matemática, obligatorio para los que desean concursar a una plaza de profesor. Sólo el 56% de ellos eran licenciados en Matemáticas o Estadística y el resto habían cursado ingeniería, arquitectura u otras ramas de ciencias. Sin embargo, todos habían realizado una o más asignaturas de estadística y más de la mitad (57%) tenía experiencia de enseñanza.

A la hora de trabajar con ellos, los futuros profesores se dividieron en dos grupos de aproximadamente el mismo tamaño, que trabajaron con el mismo profesor y método. La evaluación se llevó a cabo como parte de un taller formativo realizado dentro de una asignatura de Didáctica de la Matemática, de dos sesiones, de aproximadamente dos horas de duración. En la primera, dirigida a desarrollar y evaluar el conocimiento matemático sobre inferencia de los participantes, se resolvió un problema de contraste de hipótesis y otro de intervalo de confianza. Los futuros profesores dispusieron de una hora para resolver cada uno de los problemas, pudiendo consultar apuntes o lecciones de estadística en Internet, ya que disponían de ordenadores. La segunda parte del taller se dedicó a corregir las soluciones de los problemas y con objeto de desarrollar su conocimiento didáctico se les pidió resumir las dificultades que a su juicio podrían tener los estudiantes con los problemas propuestos. El problema analizado en este trabajo se presenta a continuación.

> 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.

Con los datos aportados, los futuros profesores debían completar todos los pasos de un contraste de hipótesis unilateral sobre la media de una población normal, cuya varianza es conocida, contenido incluido en las anteriores directrices curriculares para el Bachillerato de Ciencias Sociales (MECD, 2007) y en el temario de la oposición de profesores de Matemáticas para Educación Secundaria Obligatoria y Bachillerato. Además, en las pruebas de acceso a la universidad en Andalucía se ha incluido con frecuencia un problema similar al planteado.

Mediante la respuesta a esta tarea se pretendía evaluar en qué medida los futuros profesores tenían un conocimiento matemático suficiente del tema. En concreto, se evaluarían los tres puntos siguientes: a) Forma en que establecen las hipótesis del contraste; b) Si el procedimiento seguido para resolver el problema es consistente con las hipótesis planteadas; c) Desarrollo del procedimiento y posibles errores y d) Toma de decisión sobre la hipótesis y contextualización de los resultados.

Análisis y resultados

Nuestra investigación siguió una metodología cualitativa y exploratoria, utilizando el análisis de contenido, que busca investigar sobre la naturaleza del discurso y se utiliza para el estudio sistemático de documentos escritos (Raigada, 2002). Recogidas las soluciones detalladas al problema, se llevó a cabo un análisis cualitativo de su contenido, seleccionando las unidades del informe entregado por cada estudiante que correspondan a cada una de las variables que se analizan en el trabajo. Para cada una de estas variables se definieron categorías de análisis, partiendo de trabajos previos y creando categorías nuevas cuando fue necesario. Dichas categorías y su codificación correspondiente se revisaron mediante un proceso cíclico e inductivo típico de la investigación cualitativa. Finalmente se realizaron tablas de frecuencia para resumir los datos, como mostramos a continuación.

Planteamiento de las hipótesis

El primer paso para desarrollar un contraste de hipótesis es establecer las hipótesis nula y alternativa, cuyo planteamiento incorrecto puede derivar en la elección de un contraste inadecuado, error en el cálculo de las regiones críticas y de aceptación y, finalmente, en tomar una decisión errónea sobre el rechazo o no rechazo de la hipótesis nula (Vera et al., 2011). Teniendo en cuenta esta importancia, en primer lugar se analizaron las hipótesis planteadas por los futuros profesores, que se clasificaron según las categorías identificadas en Vera et al. (2011), que son las siguientes:

C1.1. Planteamiento correcto. Supone reconocer que el problema propuesto es un contraste sobre el valor medio de una población con distribución normal, donde la desviación típica es conocida. 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. Dado que la hipótesis nula de un contraste es la hipótesis que se desea rechazar, 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 la hipótesis alternativa será la contraria. Un ejemplo de planteamiento correcto es la respuesta del estudiante ACG, donde se observa que ambas hipótesis son complementarias y cubren el espacio paramétrico:

$$H_0{:}\mu_e \leq 69.2;\ H_1{:}\mu_e > 69.2 \text{ (Respuesta de ACG)}$$

C1.2. Plantea correctamente las hipótesis pero no utiliza la simbolización. En esta categoría se han considerado aquellos sujetos que identifican correctamente que se trata de un contraste sobre la media de una población y discriminan adecuadamente la hipótesis nula y alternativa, que plantean en términos del parámetro, son complementarias, y corresponden a un contraste unilateral. Pero, en lugar de utilizar la notación adecuada, posiblemente por no recordarla, describen las hipótesis mediante uso un lenguaje coloquial:

> 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).

C1.3. Plantea correctamente las hipótesis con simbolización incorrecta. Algunos futuros profesores llevan a cabo correctamente todos los pasos descritos en la categoría *C1.1*, sin embargo, hacen uso de una notación no empleada habitualmente, como ocurre en el caso siguiente, que no deja claro sise diferencia correctamente la media muestral y poblacional o existe una confusión implícita. El uso de una simbolización incorrecta para plantear las hipótesis también fue descrito en el estudio de Cañadas, Batanero, Díaz y Roa (2012).

Llamamos M_e a la esperanza media de vida en Europa.

$$H_0\text{:}M_e \leq 69.2 \text{ y } H_1\text{:}M_e > 69.2 \text{ (Respuesta de MTF)}$$

C1.4. Las hipótesis no cubren el espacio paramétrico. Por tanto, se trata de hipótesis no complementarias. En el siguiente ejemplo, se utiliza en el planteamiento de la hipótesis nula el símbolo $<$, 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).

Planteamos el contraste de hipótesis donde se establece la hipótesis nula, H_0 (como comprobaciónpor contraste) y la hipótesis alternativa, H_1.

$$H_0\text{:}\mu_E < \mu \rightarrow \mu_E < 69{,}2;\ H_1\text{:}\mu_E > \mu \rightarrow \mu_E > 69.2 \text{ (Respuesta de LUG)}$$

C1.5. Confunde las hipótesis nula y alternativa. Cuando el estudiante ha identificado de los datos del problema la media supuesta de la población (parámetro), por lo que queda claro que diferencia entre parámetro y estadístico (media muestral) y también deduce, a partir del enunciado, que se trata de realizar un contraste unilateral; sin embargo, intercambia las hipótesis nula y alternativa.

> Dada una población sobre la que se observa la variable X, tal que X: esperanza de vida $\rightarrow$ N(69.2.10). Tenemos $(X_1, \ldots, X_{16})$ 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.
>
> $$H_0\text{:}\mu \geq 69{,}2;\ H_1\text{:}\mu < 69{,}2 \text{ (Respuesta de AGB).}$$

La confusión entre los papeles de las hipótesis nula y alternativa fue descrita por Vallecillos (1999) y Vera et al. (2011) y se puede explicar por una falta de comprensión de la lógica del contraste de hipótesis. Esta lógica es más similar (aunque no idéntica) a la prueba por contradicción, 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. Hawkins, Jolliffe y Glickman (1991) sugieren que los estudiantes buscan resultados confirmatorios de las hipótesis estadísticas; por ello esperan que las hipótesis deban ser aceptadas y no rechazadas.

C1.6. Confunde las hipótesis nula y alternativa y la media muestral y poblacional. Al error 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, error citado en Harradine et al. (2011) 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 el estudiante explícitamente indica que la media de la muestra es 69,2 (siendo en realidad la de la población) y plantea su hipótesis utilizando dicha muestra.

$\bar{x} = 69{,}2$ años, $\sigma = 10$, $n = 16$, $\mu_0 = 78$, $\alpha = 0.05$. Tenemos el contraste $H_0{:}\mu \geq 78;\ H_1{:}\mu < 78$ (Respuesta de DCC)

C1.7. Sólo plantea una hipótesis. En el siguiente ejemplo, sólo se plantea la hipótesis nula y es difícil saber si se está implícitamente considerando la hipótesis alternativa. 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).

C1.8. Plantea solo una hipótesis y confunde la media de la muestra y de la población. Hemos considerado conveniente crear esta categoría donde se unen los dos de los errores anteriormente comentados, pues es el único caso en que se cometen dos errores al plantear las hipótesis, como se observa en la siguiente respuesta:

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

C1.9. Plantea 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 (como la esperanza media de vida) como discreta está implícita, y no es descrita en trabajos previos.

> 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).

C1.10. El valor de la media poblacional se incluye en la hipótesis alternativa. Dentro de esta categoría se han incluido aquellas respuestas en las que el valor hipotético del parámetro se asigna a la hipótesis alternativa. Aunque la probabilidad de un valor puntual de una variable 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. Este error no se ha descrito en los trabajos previos analizados y se muestra en la respuesta que reproducimos:

Planteamos el contraste de hipótesis:

$$\begin{cases} H_0{:}\mu_E < 69{,}2 \\ H_1{:}\mu_E \geq 69.2 \end{cases} \text{(Respuesta de CMM)}$$

C1.11. Incluye el valor hipotético en las dos hipótesis. El estudiante AGP no es consciente de que el caso de igualdad se asigna sólo a la hipótesis nula y lo incorpora en ambas

Publicaciones 48(2), 73-95. doi:10.30827/publicaciones.v48i2.8334
Batanero, C. et al. (2018). Conocimiento del contraste de hipótesis por futuros profesores

hipótesis. Por tanto, en esta situación las dos hipótesis no son excluyentes ya que $69{,}2$ es la intersección de ambas regiones.

$H_0 =$ 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).

En Tabla 1 se muestran los resultados referentes a este apartado, donde aproximadamente un 75% de los participantes plantean correctamente las hipótesis. Otra cuarta parte muestra algunos errores descritos en investigaciones previas (Espinel et al., 2007; Harradine et al., 2011; Vallecillos, 1999; Vera et al. 2011), 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 incluir el valor hipotético en la hipótesis alternativa o expresar las hipótesis en términos de proporción.

Tabla 1

Planteamiento de las hipótesis

Código	Respuesta	Frecuencia	Porcentaje
C1.1	Hipótesis y notación correcta	49	67,1
C1.2	Correcta sin usar simbolización	4	5,5
C1.3	Correcta con simbolización incorrecta	2	2,7
C1.4	Las hipótesis no cubren el espacio paramétrico	2	2,7
C1.5	Confunde las hipótesis nula y alternativa	3	4,1
C1.6	Confunde las hipótesis nula y alternativa y la media muestral y poblacional	3	4,1
C1.7	Sólo plantea una hipótesis.	2	2,7
C1.8	Sólo plantea una hipótesis y confunde la media de la muestra y la población	1	1,4
C1.9	Plantea las hipótesis en términos de proporciones	1	1,4
C1.10	El valor de la media poblacional se incluye en H_1	4	5,5
C1.11	Incluye el valor hipotético en las dos hipótesis.	2	2,7
Total		73	100

Elección del tipo de contraste

Una vez planteadas las hipótesis, el paso siguiente sería 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 va a utilizar y si dicho tipo es consistente con las hipótesis establecidas y el procedimiento utilizado. No hemos encontrado este tipo de análisis en investigaciones previas, por lo que este apartado constituye una aportación original.

En la respuesta correcta esperamos que el estudiante especifique que se trata de un contraste unilateral derecho, porque la región crítica la construimos a un solo lado del valor hipotético y además, la región crítica o de rechazo debe situarse a la derecha del valor hipotético (intervalos de valores superiores a 69,2). Se han definido las siguientes categorías:

C2.1. Especifica que se trata de un contraste unilateral derecho. Junto a la descripción consistente de las hipótesis y procedimiento planteado, el estudiante sitúa la región crítica a la derecha del valor hipotético. Ejemplo de esta categoría es la respuesta del estudiante 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.

> 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).

C2.2. Indica que se trata de un contraste unilateral izquierdo. El contraste se denomina izquierdo cuando la región crítica o de rechazo se encuentra a la izquierda o por debajo del valor hipotético. 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, ya que considera como $H_1{:}\mu < 78.$ El estudiante tipifica para encontrar el valor crítico en la distribución normal y define la región crítica como $(-\infty, -1{,}645)$.

> 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 =0,05; luego 1-a=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).

C2.3. Elige un contraste unilateral sin indicar la dirección de la región crítica y aplica un contraste unilateral derecho. Como ejemplo, se muestra la siguiente respuesta donde se usa una notación incorrecta (pues 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.

> 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 (media muestral=78 | μ=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).

C2.4. No especifica 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 contraste que está realizando. En el siguiente ejemplo, al calcular el *p*-valor se tiene en cuenta la región derecha, esto es, los valores del estadístico mayores que el observado:

> 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: $p_{0,05} = P(\bar{X} > 78) = P(Z > (78-69,2) / (10/\sqrt{16})) = P(Z > 3,52) = 1 - P(Z < 3,52) = 1 - 0.99978 = 0,00022$. 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).

C2.5. No especifica el tipo de contraste y desarrolla un contraste bilateral. En este caso la respuesta no es consistente con las hipótesis planteadas, pues el cálculo de los cuantiles $-Z_{\alpha/2}$ y $Z_{\alpha/2}$ a ambos lados de la media de la distribución normal van asociados al estudio de un contraste bilateral y una región crítica 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).

C2.6. No especifica el tipo de contraste y no desarrolla los pasos. En este caso no queda claro si el estudiante sabe realizar los pasos necesarios para la elaboración del contraste. En el ejemplo siguiente se muestra la respuesta de un estudiante, que no detalla el tipo de contraste a realizar y únicamente indica que se debe rechazar 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 2 se presentan los resultados obtenidos en este apartado, donde 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. Sumando 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 2

Tipo de contraste elegido

Código	Respuesta	Frecuencia	Porcentaje
C2.1	Unilateral derecho	3	4,1
C2.2	Unilateral izquierdo	1	1,4
C2.3	Unilateral; desarrolla unilateral derecho	18	24,7
C2.4	No especifica; unilateral derecho	40	54,8
C2.5	No especifica; bilateral	8	11,0
C2.6	No especifica y no desarrolla	3	4,1
Total		73	100

Desarrollo del procedimiento

Son varias las metodologías propuestas en inferencia para contrastar una hipótesis, cada una de las cuáles condiciona el desarrollo del procedimiento a seguir (Batanero, 2000; Gigerenzer, 1993; Rivadulla, 1991). El estudio del método elegido por los participantes nos ha permitido observar que la mayoría de los futuros profesores (70%) siguió la metodología propuesta por Fisher (1971/1935) 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; Inzunsa y Jiménez, 2013).

El resto de participantes siguió otros procedimientos (analizados en Rivadulla, 1991): usar el procedimiento de Neyman calculando las regiones crítica 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 llegaron a resolverlo (2,7%). Teniendo en cuenta la metodología elegida, a continuación definimos una serie de categorías en función al desarrollo consiguiente del procedimiento:

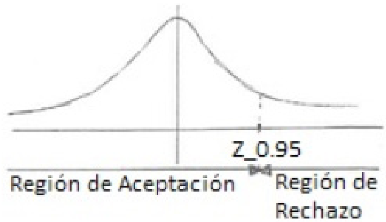
C3.1. Procedimiento correctamente desarrollado. La mayoría de los futuros profesores aplicaron correctamente la metodología de Fisher, definiendo el *p*-valor, como la probabilidad de obtener un valor de la media muestral superior o igual a 78, suponiendo la hipótesis nula H_0 cierta, es decir, asumiendo que la media poblacional toma el valor $69{,}2$. Además, dedujeron correctamente la distribución muestral, tipificando el valor obtenido del estadístico para utilizar la tabla de la distribución Normal estándar. Así, en el ejemplo que sigue, se muestra el especial cuidado en describir los pasos llevados a cabo.

$$P(\bar{X} \geq 78 | \mu = 69{,}2) = (Tipifico) = P\left(\frac{\bar{X} - 69{,}2}{2{,}5} \geq \frac{78 - 69{,}2}{2{,}5}\right) = P(Z \geq 3{,}52) = 1 - P(Z \leq 3{,}52) = (Usando\ la\ tabla\ de\ la\ N(0{,}1)) = 1 - 0{,}99978 = 0{,}00022$$

. (Respuesta de EHB).

Ejemplo de una aplicación correcta del procedimiento de Neyman es la respuesta dada por el estudiante NSF, realizando el cálculo correcto del estadístico de contraste ($Z = (\bar{X} - 69{,}2)/2{,}5 = (78 - 69{,}2)/2{,}5 = 3{,}52$) y de la región de rechazo.

> 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 $Z = (\bar{X} - 69{,}2)/2{,}5$ que sigue una distribución $\mathrm{N}\ (0{,}1)$. Por lo tanto, el valor observado del estadístico a partir de nuestra muestra es $Z_0 = (78 - 69{,}2)/2{,}5 = 3{,}52$. Como $Z_0 > z_{0,95} = 1{,}65$ entonces rechazamos la hipótesis nula para un nivel de significación $\alpha = 0.05$.



(Respuesta de NSF)

C3.2. Procedimiento correcto para un contraste bilateral. Realizar un contraste bilateral, incluso correctamente, implica un error en el cálculo del *p*-valor, región de rechazo, intervalo de confianza o función de verosimilitud, dependiendo del procedimiento seleccionado. La confusión entre el contraste unilateral y bilateral fue descrita por Espinel et al. (2007) y Vera et al. (2011). Como ejemplo, presentamos una respuesta donde se desarrolla el contraste mediante un intervalo de confianza del 95%, que es bilateral, en vez de unilateral, como sería lo correcto. Por tanto, 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 sea diferente a 69,2:

> 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).

C3.3. Procedimiento correcto usando la distribución t de Student en lugar de la normal. En este caso no se recuerdan los criterios para los que la aproximación de la distribución muestral de la media mediante la distribución normal sea adecuada y por ello se utiliza la distribución *t* de Student, con lo que el valor crítico utilizado (1,753) es incorrecto.

> 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. (Repuesta de ILB).

C3.4. Expresa el p-valor como probabilidad simple aunque lo calcula como probabilidad condicional. Algunos futuros profesores definen el p-valor como 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.

$$P(\bar{X} \geq 78) = P\left(\frac{\bar{X}-69{,}2}{2{,}5} \geq \frac{78-69{,}2}{2{,}5}\right) = P\left(Z \geq \frac{8{,}8}{2{,}5}\right) = P(Z \geq 3{,}5).$$

Y esto es lo mismo que $P(Z \geq 3{,}5) = 1 - P(Z < 3{,}5)$. (Respuesta de AMC).

C3.5. No considera explícitamente el valor observado del estadístico en el cálculo del p-valor. Otros participantes en el estudio definen el *p*-valor como 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:
>
> $$P(\bar{x} > 78) = P\left(\frac{\bar{x}-69{,}2}{2{,}5} > \frac{78-69{,}2}{2{,}5}\right) = P\left(Z > \frac{8{,}8}{2{,}5}\right) = P(Z > 3{,}5) = 0{,}00022 \quad .$$
>
> (Respuesta de IPM).

C3.6. Define el p-valor como la probabilidad de que el estadístico tome un valor puntual. En esta categoría se han incluido aquellos futuros profesores que describen el *p*-valor como la probabilidad de que el estadístico tome el valor 78, en la hipótesis establecida, es decir, $P(\bar{X} = 78 \mid \mu = 69{,}2)$ (también 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. Recordemos que si X es 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 $P(\bar{X} = 78 \mid \mu = 69{,}2)$ 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 l hipótesis nula calculando el caso más desfavorable: $P(mediamuestral = 78 \mid \mu = 69{,}2)$.(Repuesta de GPF).

C3.7. Cambia las regiones de aceptación y crítica al utilizar el método de máxima verosimilitud. El error de intercambiar las regiones 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 futuros profesores resuelven el problema mediante este método, pero en algunos casos interpretan incorrectamente los resultados, por lo que llegan a una conclusión incorrecta:

> 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:
>
> $$\varphi(X_1, \dots, X_{16}) = \begin{cases} 1 \; si \; \dfrac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} < z_{1-\alpha} \\ 0 \; si \; \dfrac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} \geq z_{1-\alpha} \end{cases}$$
>
> 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.
>
> $$\left|\frac{\bar{X}-\mu_0}{\sigma_0 / \sqrt{n}} = \frac{78-69{,}2}{10 / \sqrt{16}} = \frac{8{,}8}{2{,}5} = 3{,}52\right.$$
>
> 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 3 recoge las frecuencias y porcentajes de las categorías resultantes en el procedimiento desarrollado por los futuros profesores, 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 (61,7%), el cual es interpretado como probabilidad simple por un 11% de participantes, como probabilidad puntual por un 39,7% o el valor hipotético no es incluido en el cálculo (11%). Estos errores de interpretación del *p*-valor no han sido descritos en la investigación previa. Además, encontramos la confusión entre contraste unilateral y bilateral descrita por Espinel et al. (2007) y Vera et al. (2011).

Tabla 3

Desarrollo del procedimiento

Código	Desarrollo del procedimiento	Frecuencia	Porcentaje
C3.1	Procedimiento correcto	11	15,1
C3.2	Procedimiento correcto para contraste bilateral	13	17,8
C3.3	Procedimiento correcto usando la t de Student	1	1,4
C3.4	Expresa el *p*-valor como probabilidad simple	8	11,0
C3.5	No incluye el valor hipotético en el *p*- valor	8	11,0
C3.6	Interpreta el *p*-valor como probabilidad puntual	29	39,7
C3.7	Máxima verosimilitud, pero intercambia regiones	3	4,1
Total		73	100,0

Determinación del *p*-valor

Puesto que un alto número (51) de participantes recurren al 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 categorías:

C4.1. Cálculo del p-valor mediante tipificación y uso de la tabla de la normal estandarizada. Estos futuros profesores 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=(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).

C4.2. Aproxima 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. Lo observamos en el siguiente ejemplo que, además, define el *p*-valor como probabilidad puntual y mezcla probabilidades simples y condicionales:

$$\mathrm{P}(\overline{\mathrm{X}} = 78|\mu = 69{,}2) = \mathrm{P}\left(\frac{78 - 69.2}{2{,}5}\middle|\mu = 69{,}2\right) = \mathrm{P}(\mathrm{Z} \geq 3{,}5) = 0 < 0{,}05$$

> (Respuesta de MAM).

C4.3. Aproxima el p-valor utilizando la propiedad del porcentaje de casos incluidos en los intervalos centrales en una distribución normal. Algunos futuros profesores tratan de paliar su olvido de la lectura de las tablas recordando las propiedades de simetría de la distribución normal respecto al valor medio μ , y 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 se hace uso de la mencionada propiedad:

> 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)=0,99 (Respuesta de AMC).

C4.4. Describe 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 en que se calcularía, como vemos a continuación:

> ¿Cuál es la $P(X$=78 μ_2=69,2)? Tipificamos P = (78-69,2)/2,5; μ_2=69,2); P(Z≥3,5) (Respuesta de AGP).

El análisis de la forma en que los 51 estudiantes se basaron en el cálculo del *p*-valor llegan a determinarlo, nos ha permitido ver que el 70,6% de los mismos tipifican y realizan un uso correcto de la tabla de la distribución normal estándar. El resto presenta ciertas dificultades y bien aproximan a cero la probabilidad (7,8%), recurren a la propiedad de los intervalos centrales de una distribución normal (17,6%) y dos estudiantes (3,9%) 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. Por el contrario, en nuestro estudio no ha aparecido el intercambio de los miembros de la probabilidad condicional en la interpretación del *p*-valor encontrado en investigaciones como las de Birnbaum (1982), Falk (1986), Haller y Krauss (2002) o Vallecillos (1994).

Tabla 4

Obtención del p-valor

Códigos	Método de obtención	Frecuencia	Porcentaje
C4.1	Correcto	36	70,6
C4.2	Aproxima por cero	4	7,8
C4.3	Intervalos centrales en la normal	9	17,6
C4.4	No calcula	2	3,9
Total		51	100,0

Interpretación de los resultados

Una vez finalizado los cálculos, esperamos 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 la misma o menor que en el conjunto de países.

En el problema planteado, los futuros profesores han de recorrer un ciclo completo de modelización, descrito por Henry (1997): en primer lugar, han de abstraer la realidad (esperanza de vida y los datos del problema) para plantear 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 el paso final es traducir dicha conclusión a la realidad estudiada (esperanza de vida media al nacer). Considerando tanto la decisión tomada como la contextualización de la misma se han definido las siguientes categorías.

C5.1. Decisión correcta y contextualización adecuada. Engloba aquellos casos en los que el futuro profesor toma la decisión correcta, rechazando la hipótesis nula y argumentando 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 α=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).

C5.2. Decisión y contextualización correcta, pero interpreta el resultado en términos de probabilidad de la hipótesis. Algunos futuros profesores toman la decisión correcta, sin embargo, realizan una interpretación errónea del *p*-valor, ya que lo consideran como la probabilidad de que la hipótesis finalmente aceptada sea correcta. Estos futuros profesores estarían dando una interpretación bayesiana a los resultados del contraste, lo cual es inadecuado y fue descrito, entre otros, por Birnbaum (1982) y Falk (1986). Podemos observar este error en el ejemplo mostrado a continuación:

> 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).

C5.3. Decisión y contextualización correctas, pero 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. Dicha concepción errónea, descrita por Vallecillos (1994) e Inzunsa y Valdés (2013), 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 $(78\text{-}69{,}2)/2{,}5=3{,}52$ 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).

C5.4. Decisión correcta con contextualización inapropiada. Cuando se indica que la hipótesis nula debe ser rechazada pero dicha decisión no va en consonancia con la conclusión contextualizada. 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).

C5.5. Toma la decisión correcta, pero no contextualiza. Se han considerado aquellos casos en los que se concluye el ejercicio indicando el rechazo de la hipótesis pero sin asociar dicha respuesta con el contexto en el que se desarrolla el ejercicio:

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).

C5.6. Indica cómo tomaría la decisión y contextualiza. Este caso suele corresponder a futuros profesores que no han finalizado el procedimiento. Así, en el caso siguiente, que optó por el procedimiento de intervalo de confianza, no llega a calcular los valores numéricos del extremo, pero la contextualización es 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).

C5.7. Indica cómo tomaría la decisión, pero no contextualiza. 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, no se llega a calcular los valores numéricos del extremo, pero describe correctamente el método.

> 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 contrate hipótesis, podremos determinar si rechazamos la hipótesis nula o la retenemos. (Respuesta de MMR).

C5.8. Decisión incorrecta. A continuación incluimos una respuesta que basa su procedimiento en el intervalo de confianza y, en dicho caso, si el valor supuesto del parámetro no está incluido en el intervalo construido, la hipótesis nula debe ser rechazada. 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).

C5.9. No toma decisión. No se indica cómo y cuál es la decisión del contraste, como el caso que sigue, donde no se refleja la decisión a tomar:

> Tipificamos $P = [(78\text{-}69,2)/2,5]\ \mu_2 = 69,2\ P(Z \geq 3,5)$. (Respuesta de AGP).

En la Tabla 5 se presentan los resultados de la interpretación realizada por los futuros profesores, donde menos del 40% toman la decisión correcta, contextualizando al problema planteado; aunque otro 47,9% (categorías *C2, C3, C4* y *C5*) 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 significado de rechazar la hipótesis nula con el problema que se está tratando, es decir, de desarrollar el último paso del proceso de modelización descrito por Henry (1997). 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 5

Frecuencia y porcentaje de participantes según interpretación de los resultados

Código	Interpretación de los resultados	Frecuencia	Porcentaje
C5.1	Correcta y contextualiza	28	38,4
C5.2	Correcta, pero interpreta *p* como probabilidad de la hipótesis	3	4,1
C5.3	Correcta, contextualiza, pero indica que la hipótesis nula es falsa	2	2,7
C5.4	Correcta, contextualización no apropiada	6	8,2
C5.5	Correcta, no contextualiza	24	32,9
C5.6	Indica cómo tomaría la decisión y contextualiza	1	1,4
C5.7	Indica cómo tomaría la decisión; no contextualiza	7	9,6
C5.8	Incorrecta	1	1,4
C5.9	No toma la decisión	1	1,4
Total		73	100,0

Discusión y conclusiones

Lo más sencillo en nuestro estudio fue el planteamiento de las hipótesis, ya que la mayoría de los participantes las formularon correctamente, al igual que ocurrió en la investigación de Batanero, Vera y Díaz (2012) con estudiantes de psicología. También en su mayoría eligen un procedimiento consistente con las hipótesis planteadas, aunque muchos no declaran explícitamente el tipo de contraste elegido y otros confunden el contraste. Son pocos los que desarrollan correctamente todos los pasos del contraste elegido y algunos cometen también errores o imprecisiones en el cálculo del *p*-valor.

Los errores detectados en este trabajo reproducen los identificados en estudiantes en investigaciones previas, como las de Cañadas et al. (2012); Vallecillos (1994) y Vera et al. (2011), 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.

Por otro lado, aunque la mayoría tomó una decisión correcta en el contraste planteado, fueron pocos los que llegaron a interpretar este resultado en el contexto del problema. Este último resultado es preocupante, pues se espera que el futuro profesor desarrolle en sus estudiantes la capacidad de interpretar estudios estadísticos. Así, por ejemplo, uno de los estándares de aprendizaje evaluables para el cuarto curso de la Educación Secundaria Obligatoria indica (MECD, 2015, p. 398 y p. 407): *Interpreta un estudio estadístico a partir de situaciones concretas cercanas al alumno*. Por tanto, es necesario que un futuro profesor sea capaz de contextualizar la información resultante de un procedimiento estadístico para ayudar al alumno a adquirir esta capacidad de interpretación.

Los resultados obtenidos nos advierten de la necesidad de preparar al futuro profesorado de Educación Secundaria y Bachillerato en el tema de inferencia estadística, y más concretamente en el contraste de hipótesis. No sólo este contenido ha sido tratado en las directrices curriculares y pruebas de acceso a la universidad del Bachillerato de Ciencias Sociales en los últimos años, 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).

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

Agradecimientos

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

Referencias bibliográficas

Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores.* Tesis doctoral. Universidad de Granada.

Badenes-Ribera, L., Frías-Navarro, D., Monterde-i-Bort, H., & Pascual-Soler, M. (2015). Interpretation of the p-value: A national survey study in academic psychologist from Spain. *Psicothema, 27,* 290-295.

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning, 2*(1-2), 75-98.

Batanero, C., Vera, O. D., & Díaz, C. (2012). Dificultades de estudiantes de Psicología en la comprensión del contraste de hipótesis. *Números, 80*, 91-101.

Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics, 4*, 24–27.

Cañadas, G., Batanero, C., Díaz, C., & Roa, R. (2012). Psychology students' understanding of the Chi-squared test. *Statistique et Enseignement, 3*(1), 3-18.

Caperos, J. M., & Pardo, A. (2013). Consistency errors in p-values reported in Spanish psychology journals. *Psicothema, 25,* 408-414.

Castro Sotos, A. E., Vanhoof, S., Van den Nororgate, W., & Onghena, P. (2007). Student´s misconceptions of statistical inference: A review of the empirical evidence form research on statistical education. *Educational Research Review, 2*(2), 98-113.

Dolor, J., & 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, R. M., & Ramos, C. E. (2007). Algunas alternativas para la mejora de la enseñanza de la inferencia estadística en Secundaria. *Números, 67*, 15-23.

Falk, R. (1986). Conditional probabilities: insights and difficulties. En R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292 – 297). Victoria, Canada: International Statistical Institute.

Fisher, R. A. (1971). *The design of experiments.* Edinburgh: Oliver & Boyd (trabajo original publicado en 1935).

Gigerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. En G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methdological issue*s (pp. 311-339). Hillsdale, NJ: Lawrence Erlbaum Associates.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7*(1), 1–20.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

Harradine, A., Batanero, C., & Rossman, A. (2011). Students' and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics- Challenges for teaching and teacher education. A Joint ICMI/IASE Study.* New York: Springer.

Hawkins, A., Jolliffe, F., & Glickman, L. (1991). *Teaching statistical concepts.* London: Longman.

Henry, M. (1997). Notion de modéle et modélization en l'enseignement. En M. Henry (Ed.), *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM.

Inzunsa, S., & Jiménez, J. V. (2013). Caracterización del razonamiento estadístico de futuros profesores universitarios acerca de las pruebas de hipótesis*. Revista Latinoamericana de Investigación en Matemática Educativa, 16*(2), 179-211.

Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies, 4*(2), 126-138.

López-Martín, M. M., Batanero, C., Díaz-Batanero, C., & Gea, M. (2016). La inferencia estadística en las Pruebas de Acceso a la Universidad en Andalucía. *Revista Paranaense de Educação Matemática, 5*(8), 33-59*.*

McLean, A. (2002). Statistacy: Vocabulary and hypothesis testing. En B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching Statistics*. Cape Town: International Statistical Institute.

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 (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.* Madrid: Autor.

Morrison, D. E., & Henkel, R. E. (Eds.) (2006). *The significance tests controversy: A reader*. Chicago: Aldine, 2ª edición.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, *5*(2), 241.

Raigada, J. L. P. (2002). Epistemología, metodología y técnicas del análisis de contenido. *Sociolinguistic Studies, 3*(1), 1-42.

Ramos, C. E., Espinel, M.C., & Ramos. R. M. (2009). Identificación de los errores en los contrastes de hipótesis de los alumnos de Bachillerato. *SUMA, 61,* 35-44.

Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropos.

Vallecillos, A. (1994). *Estudio teórico-experimental de errores y concepciones sobre el contraste estadístico de hipótesis en estudiantes universitarios.*Tesis Doctoral. Universidad de Granada.

Vallecillos, A, (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute, 58,*201-204.

Vera. O., Díaz, C., & Batanero, C. (2011). Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. *Unión, 27*, 41- 61.