

Dificultades de estudiantes de Psicología en la comprensión del contraste de hipótesis

Carmen Batanero (Universidad de Granada)

Osmar Darío Vera (Universidad Nacional de Quilmes)

Carmen Díaz (Universidad de Huelva)

Fecha de recepción: 5 de noviembre de 2011

Fecha de aceptación: 27 de abril de 2012

Resumen

En este trabajo presentamos un estudio de evaluación de dificultades y errores en relación a la comprensión del contraste de hipótesis en una muestra de 224 estudiantes de Psicología, después de haber finalizado un curso de Análisis de datos. Se analizan la asignación y discriminación de hipótesis, tipos de errores, nivel de significación y potencia y regla de decisión en un contraste. Nuestros resultados son mejores que los de otros estudios previos, aunque permanecen errores relacionados con la discriminación entre los tipos de error, relación entre regiones, nivel de significación, valor p y potencia. En consecuencia se sugiere una introducción más gradual a la inferencia comenzando en la educación secundaria con actividades informales de inferencia.

Palabras clave

Contraste de hipótesis. Comprensión. Evaluación de dificultades. Estudiantes de psicología

Abstract

In this paper, we study the difficulties and errors related to statistical tests in a sample of 224 psychology students, after having finished a Data Analysis course. We analyse the assigning and discrimination of hypotheses, type of errors, significance level and power, as well as the decision rule in a test. Our results outperform those in other previous studies; although errors related to discrimination between type of error, relationships between regions, significance level, p-value and power. Consequently we suggest a more gradual introduction to inference, starting in secondary education with informal inference activities.

Keywords

Hypothesis tests, Understanding, Assessing difficulties, Psychology students.

1. Introducción

La inferencia estadística juega un papel destacado en diversas ciencias humanas, entre otras la Psicología, que basa sus investigaciones en datos recogidos en muestras de poblaciones, a las que necesita extender sus conclusiones. Sin embargo, el uso e interpretación de la estadística para las publicaciones en Psicología no es siempre adecuado, como se muestra en diversas revisiones (por ejemplo, en Harlow, Mulaik y Steiger, 1997; Ares, 1999; Borges, San Luis, Sánchez, y Cañadas, 2001; Batanero y Díaz, 2006).



Estos errores también se producen en estudiantes universitarios, como muestran muchas investigaciones (por ejemplo, Birnbaum, 1982; Vallecillos, 1994 o Krauss y Wassner, 2002), la mayoría de las cuáles se centra en la comprensión del concepto de nivel de significación. Sin embargo son muchos los conceptos que se deben comprender para llevar a cabo con éxito un contraste de hipótesis. Más aún, como se indica en Harradine, Batanero y Rossman (2011), la comprensión de la inferencia estadística requiere el aprendizaje de tres elementos interrelacionados: (a) el proceso de razonamiento; (b) los conceptos asociados y (c) los cálculos. Pero, mientras la realización de los cálculos asociados con los contraste de hipótesis es hoy día muy sencilla, gracias al software estadístico, la enseñanza de los conceptos y el razonamiento inferencial es mucho más compleja, lo que explica las muchas dificultades descritas en el uso de la inferencia.

En este trabajo abordamos un estudio de evaluación de la comprensión del contraste de hipótesis en un grupo de estudiantes de Psicología españoles, después de haber seguido un curso de inferencia. Más concretamente, a partir de sus respuestas a un pequeño cuestionario, evaluamos su comprensión de los puntos siguientes: diferencia entre contraste unilateral y bilateral, hipótesis nula y alternativa, tipos de errores y sus probabilidades, y regla de decisión en un contraste. Se ha realizado el trabajo con estudiantes de Psicología, puesto que la enseñanza de la estadística a los mismos plantea especiales problemas didácticos, al no poseer una base matemática tan amplia como la de estudiantes de otras carreras científicas. Al mismo tiempo, la inferencia estadística, resulta un instrumento importante en la investigación en Psicología, y su necesidad e importancia han ido aumentando durante los últimos años.

En lo que sigue se describen los antecedentes, muestra y método empleados y se discuten los resultados. Se finaliza con algunas reflexiones sobre las implicaciones para la enseñanza.

2. Antecedentes

Como se ha indicado, el contraste de hipótesis es una herramienta básica en investigación empírica, tanto en la confirmación de los modelos empleados, como en la demostración de la existencia de efectos de las variables de interés en la investigación (Batanero, 2000). Debido a la variedad de conceptos que deben ser comprendidos y relacionados para realizar las pruebas de hipótesis, que se confunden entre sí, el mal uso y los errores conceptuales han sido discutidos por más de 20 años, y es posible encontrar en muchas publicaciones un resumen de estas malas prácticas, que se pueden hallar incluso aún entre investigadores (Chow, 1996; Harlow, Mulaik y Steiger, 1997; Ares, 1999; Batanero, 2000; Díaz y de la Fuente, 2004; Batanero y Díaz, 2006; Castro, Vanhoof, Van den Nororgate y Onghena, 2007 y Díaz, Batanero y Wilhelmi, 2008).

2.1. Planteamiento de las hipótesis

Respecto al concepto de hipótesis, un primer error es confundir la hipótesis de investigación con una de las hipótesis estadísticas (bien nula o alternativa). Chow (1996) analiza la diferencia entre ellas: mientras que la hipótesis de investigación suele ser amplia y referirse a un constructo inobservable (por ejemplo, que el rendimiento de los estudiantes es diferente en dos poblaciones de alumnos), la hipótesis estadística, hace referencia a un parámetro de la distribución de una variable en una población de sujetos (en el ejemplo, la diferencia de medias en la puntuación en un cuestionario de dos muestras de sujetos de las citadas poblaciones). A pesar de esta diferencia, cuando encuentran un resultado significativo, algunos investigadores piensan que el resultado se puede extrapolar directamente a la hipótesis de investigación (Chow, 1996). En el ejemplo, supondría que si se rechaza que la puntuación media de las dos muestras de estudiantes en la prueba es igual, se podría deducir que el rendimiento (en general) de una de las poblaciones es mayor que el del otro. Esto podría no ser cierto, puesto que la prueba usada de rendimiento podría favorecer a uno de los grupos y si se utilizase

otra prueba diferente los resultados podrían variar.

Otro error frecuente entre alumnos e investigadores es la confusión entre las hipótesis nula y alternativa. En el ejemplo, la hipótesis alternativa sería que la media de la distribución de puntuaciones en una prueba de rendimiento de uno de los grupos será mayor que la del otro. La hipótesis nula o de no efecto es la negación de la hipótesis alternativa (en el ejemplo, que la puntuación media sería igual en los dos grupos). A pesar de esta diferencia Vallecillos (1994), en una muestra de estudiantes de distintas especialidades ($n=436$), de los cuales 70 eran estudiantes de psicología, encontró un 13% de alumnos que confundieron la hipótesis nula con la alternativa. Otro 20% de estudiantes no parecieron comprender que la hipótesis del contraste se refiere al parámetro de la población, planteando sus hipótesis utilizando el estadístico muestral. Este error, sin embargo, no fue tan frecuente en el estudio de Vera, Díaz y Batanero (2011), quienes analizan la forma en que los estudiantes de psicología plantean las hipótesis estadísticas en un problema abierto. Sin embargo, los autores encuentran alumnos que plantean hipótesis alternativas puntuales o hipótesis que en su conjunto no cubren el espacio paramétrico, de modo que coinciden con Vallecillos en que los alumnos confunden algunas propiedades de las hipótesis nula y alternativa.

2.2. Tipos de error y sus probabilidades

En cuanto a los errores en el contraste de hipótesis, la mayoría de las investigaciones se centra en la comprensión del *nivel de significación* y el *p-valor*. Batanero, Díaz y Wilhelmi (2008), indican que la interpretación incorrecta más extendida de estos conceptos es cambiar los términos de la probabilidad condicional en su definición, interpretando el nivel de significación como la probabilidad de que la hipótesis nula sea cierta, habiendo tomado la decisión de rechazarla. El mismo intercambio de condicional se hace en la interpretación del p-valor (probabilidad de obtener un valor igual o más extremo al dado, si la hipótesis es cierta) que se interpreta como probabilidad de que la hipótesis sea cierta si se obtuvo el valor dado del estadístico. Estos errores son descritos, entre otros, por Birnbaum (1982), Vallecillos (1994), Lecoutre (1999), Lecoutre, Lecoutre y Poitevineau (2001) y Haller y Kraus (2002).

A estas interpretaciones incorrectas se une, normalmente, la confusión entre significación estadística y significación práctica, que implica significación estadística más un efecto experimental (diferencia del valor del parámetro en función de una cierta variable experimental) suficientemente elevado. Aunque la significación práctica, si el tamaño de muestra es suficiente, suele ir unida a la significación estadística, sin embargo podemos encontrar datos estadísticamente significativos con un pequeño efecto experimental, siempre que tomemos una muestra grande. Esto no es siempre comprendido, como muestra Lecoutre (1999) en un experimento con 20 investigadores en Psicología y 25 profesionales estadísticos, y en el que la mayoría tenía una confianza excesiva en los contrastes y olvidaban el tamaño de la muestra.

Otro error es pensar que el p-valor indica la probabilidad de que el valor obtenido del estadístico se deba al azar, aunque esto no es cierto en general (Batanero, 2000). Por ejemplo, una diferencia significativa de las medias de un grupo experimental y otro de control puede ser debida a un tratamiento particular, pero también un grupo puede estar formado por individuos más inteligentes o con mejores medios. Moses (1992) indica otra confusión que consiste en creer en la conservación del nivel de significación cuando se realizan contrastes consecutivos en el mismo conjunto de datos.

2.3. Regla de decisión

Otro aspecto a reseñar (Vallecillos, 1994, 1999) es que los estudiantes pueden confundir *el criterio de decisión* al aplicar el contraste y por tanto, llegar a una decisión equivocada. La autora



observa una confusión entre región de rechazo y de aceptación, así como sobre la forma en que se deben construir dichas regiones en un contraste unilateral ó bilateral. Hay una falta de apreciación de que la hipótesis alternativa determina, junto con el nivel de significación, la región crítica, y que un mayor nivel de significación da una menor área para la región crítica. La comprensión de este punto también se relaciona con la dificultad señalada por Harradine, Batanero y Rossman (2011) de comprensión de la lógica subyacente al contraste de hipótesis.

En resumen, son muchas las investigaciones sobre el contraste de hipótesis. Sin embargo son pocos los trabajos en que se ha hecho un seguimiento directo de la enseñanza recibida por los estudiantes o en los que los ítems propuestos formen parte de la prueba final de evaluación de sus conocimientos. Es posible, por ello, que algunos de los errores denunciados se deban simplemente al olvido y no aparezcan con tanta frecuencia con un mayor estudio del tema por parte de los estudiantes. Con objeto de indagar en esta posibilidad, en nuestro trabajo se hizo un seguimiento de la enseñanza recibida por los estudiantes durante los dos cursos de Análisis de datos, en colaboración con el profesor de la materia. Los estudiantes fueron animados a estudiar el tema, y las preguntas formaron parte de su evaluación final. Seguidamente describimos la metodología y discutimos los resultados obtenidos.

3. Metodología de la investigación

La muestra estuvo formada por un total de 224 alumnos de segundo año de la Licenciatura en Psicología en la Universidad de Huelva, que cursaban una asignatura de Análisis de Datos II, en el segundo curso de estudios. El primer tema de la asignatura fue el contraste de hipótesis, estudiándose a continuación, los contrastes de medias, varianzas y proporciones en una y dos muestras y el análisis de varianza elemental. Los estudiantes previamente siguieron un curso de Análisis de Datos I en el primer año de sus estudios, donde estudiaron estadística descriptiva, probabilidad e introducción a la inferencia. Los dos cursos incluyeron actividades prácticas de análisis de datos, utilizando el software SPSS. Los alumnos también realizaron proyectos estadísticos, con datos recogidos por toda la clase; en dichos proyectos tuvieron que plantear una serie de hipótesis a comprobar, elegir la prueba estadística más acorde para la resolución del problema planteado, y extraer sus conclusiones, todo ello con la guía del profesor.

El cuestionario usado estuvo formado por seis ítems de opción múltiple (tres opciones por ítem), eligiendo este formato porque los alumnos estaban acostumbrados al mismo, al ser habitual en sus evaluaciones. Para elaborar el cuestionario en forma rigurosa, se comenzó con una definición semántica del constructo “comprensión del contraste de hipótesis”, delimitando las unidades de contenido que se deseaba evaluar que se recogen en la Tabla 1, con indicación del ítem que evalúa cada una de estas unidades. Los ítems concretos (Figura 1) fueron seleccionados a partir de un banco de ítems previamente construido, mediante pruebas piloto de ítems y valoración mediante el juicio de expertos.

		ÍTEMS					
		I1	I2	I3	I4	I5	I6
Contraste de Hipótesis	Contraste bilateral y unilateral. Asignación de hipótesis	X	X			X	
	Hipótesis nula y alternativa. Asignación de hipótesis	X	X	X	X		X
	Errores tipo I y II, nivel de significación y potencia				X		
	Cálculo de valor crítico, dado el nivel de significación					X	
	Regla de decisión en un contraste						X

Tabla 1. Contenidos evaluados por ítem

El primer ítem evalúa la comprensión de la asignación de hipótesis estadísticas, específicamente

si el alumno es capaz de elegir la hipótesis nula adecuada, partiendo de un contexto de aplicación. La respuesta correcta es la (c), puesto que la hipótesis nula es la contraria a la que desea probar el investigador (Chow, 1996). El distractor (a) lo elegiría un alumno que confunde las hipótesis nula y alternativa, confusión descrita por Vallecillos (1994). El distractor (c) lo tomaría el estudiante que confunde un contraste unilateral con otro bilateral.

También el segundo ítem evalúa el planteamiento de las hipótesis estadísticas, más concretamente el conocimiento de las reglas para establecerlas. La respuesta correcta es la (b), ya que los conjuntos donde se define el parámetro para cada una de las hipótesis deben ser excluyentes y cubrir el espacio paramétrico (Vallecillos, 1994). Estas son las condiciones exigidas al plantear correctamente las hipótesis, que no se cumplen en los otros distractores; en el distractor (a) las hipótesis no son complementarias ni excluyentes; en el distractor (c) están intercambiadas las hipótesis nula y alternativa.

El ítem 3 permite evaluar la comprensión de las diferencias entre potencia de una prueba y probabilidad de error tipo II, y probabilidad de error tipo I. Vallecillos (1994) no incluye ítems sobre el concepto de potencia, aunque sí sobre el error tipo I y II. La respuesta correcta es la (b), ya que la potencia se define como la probabilidad de rechazar la hipótesis nula siendo falsa. El alumno que elige el distractor (a) confunde la potencia con la probabilidad de error tipo II, mientras quienes eligen el distractor (c) confunden la probabilidad de error tipo II con la probabilidad de no cometer dicho error.

El ítem 4 evalúa si el alumno es capaz de diferenciar entre los errores tipo I y tipo II, y la comprensión de la diferencia entre hipótesis estadísticas e hipótesis de investigación (Chow, 1996). La respuesta correcta es la (b), ya que se cometerá un error de tipo II si decide que no existe diferencia entre los dos tratamientos para curar la depresión, cuando en realidad si existen. El alumno que elige el distractor (a) no logra interpretar el enunciado, ya que el problema afirma se conoce que ambos medicamentos han sido probados, resultando efectivos. Quien elige el distractor (c) estaría confundiendo el error de tipo I con el de tipo II.

El ítem 5 evalúa la comprensión de la relación entre el nivel de significación y el valor crítico, así como la diferencia entre test unilateral y bilateral, y el concepto de puntuación crítica; también se requiere el manejo de tablas estadísticas. La respuesta correcta es la (b), pues si el nivel de significación es 0,01 para un test unilateral derecho, ese es el área que deja a la derecha el percentil 2,33 de la distribución normal. En el distractor (a) se confunde con el test unilateral izquierdo. El distractor (c) implica errores en el manejo de las tablas.

En el ítem 6 se evalúa la comprensión de la regla de decisión entre las hipótesis y las condiciones en las que se tomará la decisión de rechazar la hipótesis nula. La respuesta correcta es la (c), puesto que la hipótesis nula se rechaza tanto si se produce la situación ilustrada en (a) como en (b). El distractor a) la respuesta es verdadera, pero no es el único argumento para rechazar la hipótesis nula en un contraste, por eso quien lo elige no conoce la posibilidad que brinda el distractor b). Para el distractor (b) los estudiantes no asocian el cálculo de una probabilidad con la regla de decisión, no logran comprender ni son capaces de relacionar estos conceptos abstractos, llevándolos a errores generalizados en la aplicación (Batanero, 2000).



Ítem 1. Queremos conocer si los sujetos extrovertidos e introvertidos difieren en la puntuación media en autoestima y no disponemos de ninguna información previa. El tipo de hipótesis nula razonable que debo plantear es:

- $H_0 : \mu_I \leq \mu_E$
- $H_0 : \mu_I \geq \mu_E$
- $H_0 : \mu_I = \mu_E$**

Ítem 2. De los siguientes pares de afirmaciones, indique cual **NO** cumple con las reglas para plantear hipótesis estadísticas:

- $H_0 : \mu \leq 100; H_1 : \mu > 100$
- $H_0 : \sigma = 15; H_1 : \sigma \leq 15$**
- $H_0 : \mu_1 - \mu_2 = 0; H_1 : \mu_1 - \mu_2 \neq 0$

Ítem 3. Si en una investigación lees que la potencia de un contraste vale 0,5594, entonces interpretas que:

- La probabilidad de rechazar la hipótesis nula siendo falsa vale $(1 - 0,5594) = 0,4406$
- La probabilidad de rechazar la hipótesis nula siendo falsa vale 0,5594**
- La probabilidad de mantener la hipótesis nula siendo falsa vale 0,5594

Ítem 4. Supongamos que conocemos la ‘verdad absoluta’ sobre la eficacia de dos tratamientos (A y B), y sabemos que existen diferencias en la efectividad de ambos para curar la depresión. Un investigador que realice un estudio y parta de la hipótesis ‘no existen diferencias en la efectividad de los tratamientos A y B para curar la depresión’ cometerá un error tipo II cuando:

- Concluya que A y B no son efectivos para curar la depresión
- Concluya que A y B no difieren en su efectividad para curar la depresión**
- Concluya que A y B difieren en su efectividad para curar la depresión

Ítem 5. El valor crítico correspondiente a un $\alpha = 0,01$ en un contraste unilateral derecho, suponiendo normalidad de los datos es:

- 2,33
- 2,33**
- 3,10

Ítem 6. Cuando realizamos un contraste, la regla de decisión nos lleva a rechazar la hipótesis nula siempre que:

- El estadístico de contraste caiga en la región de rechazo
- La probabilidad asociada al estadístico de contraste (el nivel de significación) sea menor que el valor de α
- Las alternativas a) y b) son correctas**

Figura 1. Cuestionario

4. Resultados y discusión

Como se ha indicado, los estudiantes completaron los ítems como parte de una de las evaluaciones de la asignatura. Para evitar que respondieran al azar se les advirtió que se penalizaría las respuestas incorrectas. La Tabla 2 muestra los porcentajes de respuestas por ítem y por distractor, destacándose en negrita las respuestas correctas. Al contrario que lo expuesto en otras investigaciones, el porcentaje de estudiantes que dan respuestas correctas supera el 50% en todos los ítems, llegándose en algunos casi al total de la muestra. Todo ello confirma nuestra hipótesis sobre el efecto de la enseñanza en la superación de los errores descritos en los antecedentes.

Como observamos en el ítem 1, donde el 93,3% de los estudiantes fueron capaces de elegir la hipótesis nula adecuada, partiendo de un contexto de aplicación, fue fácil para los estudiantes comprender el planteamiento de las hipótesis en un contraste estadístico y la diferencia entre hipótesis estadística nula y alternativa. Nuestros resultados son mejores que los de Vallecillos (1994), quien, en un ítem similar obtuvo un 68,6% de respuestas correctas en la muestra global de su estudio, y un 78,6% en una submuestra de alumnos de Psicología. Un 1,8% confunde las hipótesis nula y alternativa (distractor a), confusión que en la investigaciones de Vallecillos (1994) se dio en 6,4% en la muestra global y 5,7% en Psicología. Hawkins, Jolliffe y Glickman (1991) sugieren que esta confusión también se debe a que los estudiantes buscan resultados confirmatorios (y no falsatorios) de las hipótesis estadísticas; por ello no comprenden que las hipótesis nulas están redactadas en términos de la no existencia de efectos. Otro 4,5% elige el distractor (b), confundiendo un contraste bilateral con uno unilateral error denunciado por Batanero (2000) y tan sólo un estudiante no responde (0,4%), mientras que en Vallecillos son 6,2% en la muestra global y 4,4% para el grupo de Psicología los que no responden. La diferencia entre nuestro ítem y el de Vallecillos es que el suyo se refiere a un contraste bilateral, mientras que el nuestro plantea uno unilateral.

Respuesta	Ítem 1	Item 2	Item 3	Item 4	Item 5	Item 6
a	1,8%	5,8%	17,4%	3,1%	23,2%	29,5%
b	4,5%	84,8%	50,9%	64,7%	64,3%	1,8%
c	93,3%	5,8%	20,1%	17,4%	4,5%	58,0%
sin contestar	0,4%	3,6%	11,6%	14,7%	8%	10,7%

Tabla 2. Porcentajes de respuestas por Ítem (n=224)

El ítem 2 referido asimismo al planteamiento de las hipótesis fue respondido correctamente por la mayoría de los estudiantes (84,8% de respuestas correctas); de nuevo se obtuvieron mejores resultados que en la investigación de Vallecillos (1994), quien encontró un 56% de respuestas correctas en la muestra global (77,1% en Psicología) en un ítem relacionado con éste, en donde pregunta, entre varias posibilidades, cuál de todas no es una hipótesis nula correcta. Nuestros resultados son bastante mejores, teniendo en cuenta que en nuestro caso, los alumnos han de valorar conjuntamente las dos hipótesis (en lugar de una sola). Los distractores (a; hipótesis que no cubren el espacio muestral) y (c; intercambio entre hipótesis nula y alternativa) tan sólo son elegidos por un 5,8% de alumnos (Vallecillos, 1994 no incluye estos distractores). Finalmente es muy bajo el porcentaje que no responde al ítem (3,6%), mientras que Vallecillos (1994) presenta un 17% de no respuesta, que en la especialidad de Psicología se reducen al 11,4%.

El ítem 3, que evalúa la comprensión de los conceptos de error tipo II y potencia, obtiene un 50,9% de respuestas correctas. Vallecillos (1994) obtiene 22,9% de respuestas correctas en la muestra global (20% en Psicología) en un ítem en que pregunta por la probabilidad de error tipo II (aunque no por la potencia) y su discriminación con la probabilidad de error tipo I. La autora no incluye ítems sobre el concepto de potencia. Un 17,4% elige el distractor (a), confundiendo potencia y nivel de significación (probabilidad de rechazar la hipótesis nula siendo verdadera). Un 20,1% de nuestros estudiantes elige el distractor (c), confundiendo potencia y error tipo II (probabilidad de aceptar la hipótesis nula siendo falsa). También hay un alto porcentaje de alumnos que no responden el ítem (11,6%). En resumen, la comprensión del concepto de potencia es difícil, en línea con lo sugerido por Batanero (2000), Castro et al. (2007) y Díaz, Batanero y Wilhelmi (2008).



El ítem 4, sobre discriminación entre los tipos de error da un 64,7% de respuestas correctas. Un 17,4% elige el distractor (c) confundiendo la definición de error tipo II con la definición de error tipo I. Vallecillos (1994) plantea dos ítems relacionados con la afirmación planteada en este distractor. En el primero, donde se pide identificar la definición correcta de error tipo I, un 6,4% de estudiantes confunden las definiciones de error tipo I y tipo II (11,4% en psicología). En el segundo ítem (que pregunta cuándo se comete error tipo I, una vez tomada una decisión) obtiene un 20% de estudiantes que dan erróneamente las condiciones para cometer error tipo II (24,3% en psicología). En resumen, nuestros estudiantes muestran menor confusión que los de Vallecillos en la definición formal de los tipos de error, pero mayor confusión en cuando a las condiciones prácticas en que se cometen cada uno de ellos. Sólo un 3,1% de estudiantes elige el distractor (a) no interpretando enunciado y 14,7% no responden.

En el ítem 5, que evalúa la comprensión de la relación entre nivel de significación y región crítica, el porcentaje de respuestas correctas es del 64,3%. Un 23,2% ha decidido escoger el distractor (a), confundiendo el contraste unilateral derecho con el izquierdo. Se confirma lo expuesto por Vallecillos (1994), quien sugirió que los estudiantes pueden confundir la manera de construir las regiones de aceptación y rechazo, cuando se trate de un test unilateral ó bilateral. Un 4,5% elige el distractor (c) mostrando dificultad en la lectura de la tabla de la distribución normal, error que también fue descrito por Tauber (2001) en su investigación sobre la construcción de significados acerca de la distribución normal. Vallecillos no incluye en su trabajo ningún ítem que evalúe estas relaciones.

Finalmente, en el ítem 6, sobre criterio de decisión en un contraste, obtuvimos un 58% de respuestas correctas. Un 1,8% de estudiantes elige como correcto el distractor (b), y un 29,5% el distractor (a); estos alumnos eligen únicamente una de las respuestas (a) o (b), a pesar que las dos son correctas. En consecuencia, el concepto de región de rechazo ha sido comprendido por la mayoría de los estudiantes, pero no todos comprenden que si el estadístico está incluido en la región de rechazo, el valor p correspondiente será menor que el nivel de significación. Dichos estudiantes muestran dificultad de comprensión de la idea de distribución muestral, uno de los conceptos más relevantes en inferencia (Harradine, Batanero y Rossman, 2011). Vallecillos (1994) plantea un ítem relacionado con éste en que se da una hipótesis nula y dos alternativas diferentes, preguntando por la hipótesis que se debe aceptar. Para ello el alumno ha de completar el contraste a partir de unos datos dados en el enunciado. La autora encuentra un 33,7% de respuestas correctas (42,9% en el grupo de estudiantes de Psicología), resultado peor que el nuestro. Sin embargo, hay que tener en cuenta que el ítem planteado por ella es más difícil, en cuanto los fallos podrían producirse por un error aritmético en el cálculo del estadístico de contraste o de lectura de las tablas de la distribución.

5. Discusión y conclusiones

En nuestro trabajo se obtienen mejores resultados en todos los ítems que en otros similares utilizados en la investigación previa y más concretamente en Vallecillos (1994), a pesar de que nuestra muestra está formada exclusivamente por estudiantes de Psicología, mientras que en la de la citada autora intervinieron estudiantes de diferentes especialidades. También sobrepasan nuestros resultados a las submuestras de estudiantes de Psicología en la investigación de Vallecillos (1994). Pensamos que estos mejores resultados se deben a que los estudiantes de nuestra investigación habían seguido dos cursos de estadística (en primer y segundo curso de Psicología) estudiando en el primero el contraste de hipótesis en forma general (los fundamentos) y teniendo en el segundo ocasión de aplicarlo a diferentes problemas, tales como comparación de medias y proporciones en una y varias muestras, así como en el análisis de la varianza.

Lo más sencillo resultó el planteamiento de las hipótesis (ítems 1 y 2) mostrando los estudiantes capacidad para traducir el enunciado de un problema a hipótesis estadísticas, diferenciar hipótesis nula

y alternativa, así como contraste unilateral y bilateral (ítem 1) y comprender que las hipótesis deben plantearse de modo que el conjunto de valores del parámetro en ambas hipótesis cubran el espacio paramétrico y sean conjuntos excluyentes (ítem 2). En consecuencia ha sido poco frecuente la confusión entre hipótesis nula y alternativa que Vallecillos (1994; 1999) encontró un 13% aproximadamente de alumnos en su estudio.

Encontramos que algunos errores continúan después de la enseñanza recibida por estos estudiantes, que como se ha indicado, fue de dos cursos de estadística y se hizo especial énfasis en el contraste de hipótesis. Respecto al concepto de potencia (tema no abordado en la investigación de Vallecillos), un 17,4% de estudiantes lo confunde con el nivel de significación y otro 20,1% con el error tipo II (ítem 3). Pensamos que estas confusiones, no señaladas en la literatura previa, son importantes pues Valera, Sánchez y Marín (2000) señalan que, entre todas las críticas a la aplicación actual del contraste de hipótesis, la más importante es el error en el cálculo de la potencia. Dicho error afectará a la validez del resultado, debido a la alta probabilidad de obtener resultados no significativos ambiguos, en los que no se sabe si la hipótesis nula podría ser cierta o el resultado se debe a una baja potencia.

Observamos en el ítem 4 un porcentaje apreciable de alumnos que, o bien, confunden los errores I y II o muestran confusión sobre su significado, confirmando lo sugerido por Batanero, Díaz y Wilhelmi (2008). Sería necesario insistir sobre el significado de estos conceptos, pues su falta de comprensión inhabilita al investigador para interpretar los resultados de sus contrastes o para leer en forma crítica los resultados de las investigaciones en revistas de su especialidad. Un 23,2% en el ítem 5 confunde la región de aceptación y rechazo para el contraste unilateral derecho y el izquierdo, confirmando lo expuesto por Vallecillos (1994), quien observa una confusión de la forma cómo hay que construir dichas regiones. Esta confusión llevará al estudiante a tomar una decisión equivocada (aceptar cuando debiera rechazar o viceversa). Finalmente, en el ítem 6 alrededor de un 29% de estudiantes no relacionan la región de aceptación /rechazo con el valor p y el nivel de significación, conceptos que debieran estar suficientemente claros y bien relacionados, para ser capaces de construir las regiones críticas y alternativa.

En resumen, por un lado en todos los ítems se obtuvieron mejores resultados que en otros similares utilizados en la investigación previa, donde no se describe la enseñanza recibida por las muestras participantes. Pero, por otro lado, no todos los errores logran erradicarse en los dos cursos seguidos por los alumnos. Estos resultados son razonables, debido al gran número de conceptos que deben discriminar y relacionar los estudiantes en un contraste de hipótesis correcto. También apoyan la sugerencia de Vera, Díaz y Batanero (2011) sobre la necesidad de revisar la enseñanza de la inferencia estadística, puesto que los errores encontrados son fundamentales en la elaboración de un contraste de hipótesis

Seguramente hay que aceptar la conclusión de Harradine, Batanero y Rossman (2011) de que el razonamiento inferencial no puede desarrollarse en un espacio corto de tiempo y sería importante comenzar a introducirlo de forma informal desde la enseñanza secundaria. Se podría ayudar también a los estudiantes en su construcción del razonamiento inferencial, comenzando con el planteamiento de actividades informales de inferencia, como propone Rossman (2008) antes de iniciar el aprendizaje formalizado de los contrastes de hipótesis. Esperamos en consecuencia que nuestros resultados animen a otros investigadores a proseguir analizando las dificultades de los estudiantes en la inferencia estadística y a proponer acciones educativas que contribuyan a mejorar el aprendizaje de estos conceptos.

Agradecimientos: Proyecto EDU2010-14947 (MCINN-FEDER) y grupo FQM126 (Junta de Andalucía).



Bibliografía

- Ares, V. M. (1999). La prueba de significación de la «hipótesis cero» en las investigaciones por encuesta. *Metodología de Encuestas*, 1, 47-68.
- Batanero, C. (2000). Controversies around significance tests. *Journal of Mathematics Thinking and Learning*, 2(1-2), 75-98.
- Batanero, C. y Díaz, C. (2006). Methodological and didactical controversies around statistical inference. *Actes du 36ièmes Journées de la Société Française de Statistique* [CD-ROM]. Paris: Société Française de Statistique.
- Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics*, 4, 24-27.
- Borges, A., San Luis, C., Sánchez, J. A. y Cañadas, I. (2001). El juicio contra la hipótesis nula: muchos testigos y una sentencia virtuosa. *Psicothema*, 13 (1), 174-178.
- Castro, A. E., Vanhoof, S., Van den Nororgate, W., Onghena, P. (2007). Student's misconceptions of statistical inference: A review of the empirical evidence from research on statistical education. *Educational Research Review*, 2 (2), 98-113.
- Chow, L. S. (1996). *Statistical significance: Rationale, validity and utility*. London: Sage.
- Díaz, C., Batanero, C. y Wilhelmi, M. R. (2008). Errores frecuentes en el análisis de datos en educación y Psicología. *Publicaciones*, 35, 109-133.
- Díaz, C. y de la Fuente, E. I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*, Volumen especial 2004, 161-167.
- Haller, H. y 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. y Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.
- Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 235-246). New York: Springer.
- Hawkins, A., Jolliffe, F. y Glickman, L. (1991). *Teaching Statistical Concepts*. London: Longman.
- Krauss, S., & Wassner, K. (2002). How significance tests should be presented to avoid the typical misinterpretations. En 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.
- Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2) (pp. 205 – 208). Helsinki: International Statistical Institute.
- Lecoutre, B., Lecoutre M. P., y Poitevineau J. (2001). Uses, abuses and misuses of significance tests in the scientific community: Won't the Bayesian choice be unavoidable? *International Statistical Review*, 69, 399-418.
- Moses, L. E. (1992). The reasoning of statistical inference. In D. C. Hoaglin y D. S. Moore (Eds.), *Perspectives on contemporary statistics* (pp. 107-122). Washington, DC: Mathematical Association of America.
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19. Online: <http://www.stat.auckland.ac.nz/serj>.
- Tauber, L. (2001). *Significado y comprensión de la distribución normal a partir de actividades de análisis de datos*. Tesis Doctoral. Universidad de Sevilla. España.
- Valera, A. Sánchez-Meca, J. y Marín, F. (2000). Contraste de hipótesis e investigación psicológica española: análisis y propuestas. *Psicothema* 12 (2), 549-552.
- Vallecillos, A. (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral. Universidad de Granada.
- Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. In *Proceedings of the 52 session of the International Statistical Institute* (pp. 201-204). Helsinki:

International Statistical Institute. Tome 58, Book 2.

Vera, O. D., Díaz, C. y Batanero, C. (2011). Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. *Unión*, 27. Recuperado de: www.fisem.org/web/union/.

Carmen Batanero, Facultad de Ciencias de la Educación, Universidad de Granada, es doctora en Matemáticas. Su línea de investigación es la Educación Estadística, tema sobre el que ha coordinado varios grupos de trabajo y dirigido tesis 14 doctorales. Fue presidenta de la International Association for Statistical Education (IASE) y miembro de la International Commission on Mathematical Instruction (ICMI). Ha publicado trabajos en revistas nacionales e internacionales (ver <http://www.ugr.es/~batanero/>).

Osmar Darío Vera, Universidad Nacional de Quilmes. Es licenciado en Matemáticas por la Universidad de Buenos Aires, Máster en Estadística por la Pontificia Universidad Católica de Chile y en Didáctica de las Matemáticas por la Universidad de Granada. Ha sido becado por la Fundación Carolina en Argentina para realizar su tesis doctoral en la Universidad de Granada. Ha publicado trabajos en revistas nacionales e internacionales

Carmen Díaz, Facultad de Ciencias de la Educación, Universidad de Huelva, es doctora en Psicología. Fue becaria del Plan de Formación del Profesorado Universitario y es actualmente Profesora Contratado Doctor en área de Metodología de las Ciencias del Comportamiento. Su línea de investigación son la enseñanza y aplicaciones de la estadística, tema en el que ha codirigido dos tesis doctorales. Ha publicado trabajos sobre inferencia estadística en revistas y congresos nacionales e internacionales

