

Dificultades de estudiantes de psicología en relación al contraste de hipótesis

Vera, Osmar¹ y Díaz, Carmen²

¹Universidad Nacional de Quilmes

²Universidad de Huelva

Resumen

El contraste de hipótesis es una herramienta fundamental en la inferencia estadística, aunque su mal uso en la investigación ha sido criticado extensamente. Debido a ese mal uso es que consideramos que la formación de los profesionales de las Ciencias Sociales en general y de la Psicología en particular, se debe profundizar en las dificultades que los estudiantes encuentran en los conceptos relacionados con el contraste de hipótesis.

En este trabajo presentamos un estudio de evaluación en relación al contraste de hipótesis, describimos las dificultades en una muestra de 93 estudiantes de segundo año de Psicología. Los estudiantes completaron los ítems como parte de la evaluación final de una asignatura, para asegurar que hubieran estudiado previamente. Los ítems evalúan conceptos tales como: asignación de hipótesis estadísticas, diferenciación entre hipótesis nula y alternativa, entre contraste unilateral y bilateral, asignación del valor p y nivel de significación, determinación de la zona de rechazo y aceptación y regla de decisión.

Los resultados son en general buenos, encontrando en todos los ítems un porcentaje de respuestas correctas superiores al 50%, salvo en el ítem que evalúa la diferenciación entre contraste unilateral derecho e izquierdo, confundiendo un 46% de los estudiantes el signo del valor crítico correspondiente para un nivel de significación específico. Otros errores encontrados están relacionados con la confusión entre hipótesis nula y alternativa, errores tipo 1 y 2 y la regla de decisión en el contraste de hipótesis. Los resultados se comparan con la investigación de Vallecillos (1994).

Palabras clave: contraste de hipótesis, dificultades, estudiantes universitarios, psicología.

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 en las publicaciones en Psicología no es siempre adecuado, como se muestra en diversas revisiones (Harlow, Mulaik y Steiger, 1997; Borges, San Luis, Sánchez, y Cañadas, 2001; Díaz y de la Fuente, 2004; Batanero y Díaz, 2006). Múltiples autores dedican su investigación al tema de la inferencia estadística y las dificultades que este tema suscita, ganando día a día gran impacto a nivel internacional (Haller & Krauss, 2002).

Estos errores también se producen en investigadores y estudiantes universitarios, como muestran muchas investigaciones (Birnbaum, 1982). En España, la investigación didáctica en el tema fue iniciada por Vallecillos con su tesis doctoral (Vallecillos, 1994) y ésta ha sentado bases de una amplia investigación posterior (Díaz, 2007; Alvarado, 2007; Olivo, 2008). La mayoría de las investigaciones llevadas a cabo hasta ahora se limitan a 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.

Vallecillos (1994) describe diversos errores relacionados con la inferencia estadística en una muestra de 436 estudiantes de magisterio, clasificándolos en varias categorías:

- *Parámetro y distribuciones muestrales*: Los estudiantes no dan cuenta de la variabilidad de la distribución muestral del estadístico, ni reconocen su relación con el parámetro poblacional.
- *Confusión entre las hipótesis*: Los estudiantes confunden la hipótesis nula con la hipótesis alternativa. Definen como hipótesis nula aquella que se desea probar. Por otro lado, aunque la teoría estadística se ocupa del último nivel de hipótesis (de las hipótesis estadísticas), alumnos e investigadores confunden los diferentes niveles de abstracción de las hipótesis y cuando encuentran un resultado significativo, lo interpretan en relación a la hipótesis de investigación (aunque el resultado solo se refiere a las hipótesis nula y alternativa) (Chow, 1996, Batanero, 2000).
- *Interpretación de las probabilidades de error y sus relaciones*: Los estudiantes confunden las dos probabilidades condicionales que intervienen en la definición del nivel de significación, interpretando α como $P(H_0 \text{ cierta} / \text{se ha rechazado } H_0)$ lo cual sería una interpretación bayesiana; en otros casos suprimen la condición en la probabilidad ó la confunden con su probabilidad complementaria. Este error fue también descrito por Birnbaum (1982) y Falk (1986), quien lo denominó “falacia de la condicional transpuesta”. Este error también se extiende a la confusión entre errores tipo I y tipo II.
- *Nivel de significación*: Se confunde el nivel de significación con la probabilidad de obtener un resultado correcto, y suelen pensar que está probabilidad está predeterminados por el tamaño de la muestra. Cohen (1994) indica que la interpretación errónea del valor- p como probabilidad de que la hipótesis nula sea falsa es casi universal y que, además, esta interpretación viene acompañada del error de suponer que su complementario es la probabilidad de que la siguiente replicación del experimento tendrá éxito.
- *Criterio de decisión*: Confusión entre región de rechazo y de aceptación.
- *Interpretación de resultados*: Confunden significación estadística con significación práctica, así como un resultado significativo lo ven como el resultado que corrobora la hipótesis nula y en particular al efecto de la variable manipulada por el investigador (Falk, 1986; Granaas, 2002).

En cuanto a los estudiantes de Psicología se refiere, la enseñanza de los conceptos ligados al contraste de hipótesis plantea especiales problemas didácticos, al no poseer una base matemática tan amplia como 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. Mientras la realización de los cálculos asociados a 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 (Brewer, 1985).

En nuestro trabajo queremos contribuir a este tema, analizando las dificultades que persisten en los estudiantes españoles de psicología con la administración de algunos ítems de evaluación.

2. Método

La muestra estuvo formada por un total de 93 alumnos de segundo año de la Licenciatura en Psicología en la Universidad de Huelva, que cursaban una asignatura de Análisis de Datos II. Esta asignatura (60 horas lectivas) se enmarca dentro del área de Metodología de las Ciencias del Comportamiento y ofrece a los alumnos un cuerpo de conocimientos sobre inferencia, regresión y análisis de varianza. Estos alumnos habían cursado el año anterior Análisis de Datos en Psicología I, que incluye la estadística descriptiva univariante y bivalente. En los dos cursos incluyeron actividades prácticas de análisis de datos, utilizando el software SPSS.

Cabe señalar también que al final del curso el alumno elabora un informe de un trabajo desarrollado en forma de una pequeña investigación. Con los datos recogidos por toda la clase, los alumnos por grupos plantean una serie de hipótesis a comprobar, eligen la prueba estadística más acorde para la resolución del problema, y extraen sus conclusiones. El profesor los guía en cuestiones relativas a la formulación de hipótesis, construcción de la base de datos, tratamiento y análisis de los mismos. Finalmente el alumno debe presentar sus resultados en forma de artículo según las normas APA. Por todo ello es que tanto la elección, lectura y análisis de resultados de las pruebas de hipótesis son fundamentales para aprobar el curso.

Sobre los conceptos descritos en la introducción nos propusimos elaborar un conjunto de ítems que reflejaran de manera exploratoria las dificultades de los alumnos de Psicología en los conceptos relacionados con la inferencia estadística. Para redactar los ítems se partieron de los ítems propuestos en la investigación de Vallecillos (1994) y Olivo (2008), de los que se procedió a revisar en redacción y contexto (se redactaron con ejemplos familiares para los estudiantes de Psicología). Los contenidos evaluados se recogen en la Tabla 1.

Tabla 1. Contenidos evaluados en el cuestionario

Contenido evaluado	I1	I2	I3	I4	I5	I6
Asignación de hipótesis	x					
Diferenciar hipótesis nula y alternativa	x					
Valor p, nivel de significación					x	
Diferenciar potencia, nivel de significación, probabilidad de error II		x	x			x
Región crítica y de aceptación				x	x	x
Contraste unilateral y bilateral				x		
Regla de decisión					x	x

Los estudiantes completaron los ítems como parte del examen final de la asignatura. Dado que no eran los únicos contenidos evaluados en la asignatura algunos ítems se redactaron para explorar más de un contenido, para no sobrecargar a los alumnos con más ítems (solo disponen de dos horas para completar el examen). Para evitar que respondieran al azar se les advirtió que se penalizaría las respuestas incorrectas. Los ítems aplicados se recogen en la Figura 1.

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:

- a) $\mu_I \leq \mu_E$ b) $\mu_I = \mu_E$ c) $\mu_I \geq \mu_E$

Ítem 2. Supongamos un contraste bilateral sobre la media, siendo la variable estudiada la inteligencia. Para $H_0: \mu = 100$, $H_1: \mu = 110$, $\alpha = 0,05$ y $\beta = 0,4406$. ¿Cuál es la

probabilidad de rechazar la H_0 cuando no es ‘cierta’?

- a) 0,05 b) 0,4406 **c) 0,5594**

Ítem 3. 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:

- a) Concluya que A y B no son efectivos para curar la depresión
b) Concluya que A y B no difieren en su efectividad para curar la depresión
c) Concluya que A y B difieren en su efectividad para curar la depresión

Ítem 4. La puntuación típica correspondiente a una $\alpha = 0,01$ en un contraste unilateral derecho es:

- a) 2,33** b) -2,33 c) 3,10

Ítem 5. ¿Cuál de las siguientes hipótesis está **bien** formulada?:

- a) $H_0: \mu = 3; H_1: \mu \neq 4$, b) $H_0: \mu = 3; H_1: \mu \geq 3$, **c) $H_0: \mu = 3; H_1: \mu \neq 3$**

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

- a) El estadístico de contraste caiga en la región de rechazo
b) La probabilidad asociada al estadístico de contraste (el valor de significación) sea menor que el valor de alfa
c) a) y b) son correctas

Figura 1. Ítems aplicados

3. Resultados y discusión

Como vemos en la Tabla 2, los ítems en general, no fueron excesivamente difíciles, excepto en el ítem 4 (para el cual no se tiene una referencia en investigaciones previas) más de la mitad de los alumnos responden correctamente. Nuestros resultados son en general mejores que los de Vallecillos (1994) quien obtiene respectivamente 68,6%, 22,9%, 22,9%, 56% y 33,7% % en su muestra global y algo más con alumnos de psicología (78,6%, 20%, 20%, 77,1% y 42,9%) en los ítems 1, 2, 3, 5 y 6. Los principales resultados encontrados son:

- En el ítem 1, que evalúa la asignación de hipótesis, es decir la traducción de hipótesis de investigación a hipótesis estadísticas, muestra un 90,9% de respuestas correctas (tabla 3), por lo que ha resultado fácil para los estudiantes, quienes parecen comprender el planteamiento de las hipótesis en un contraste estadístico. Un 4,3% de los estudiantes elige el distractor a) confundiendo las hipótesis nula y alternativa, confusión encontrada también en Vallecillos (1994) aunque sólo en un 2%.

Tabla 2. Porcentajes de respuestas por Ítem (n=93)

Respuesta	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6
a	4,3%	10,9%	1,1%	34,8%	3,3%	26,1%
b	93,5%	6,5%	65,2%	46,7%	1,1%	2,2%
c	2,2%	76,1%	18,5%	5,4%	94,6%	59,8%

sin contestar	1,1%	7,6%	16,3%	14,1%	2,2%	13,0%
---------------	------	------	-------	-------	------	-------

- En el ítem 2 un 10,9% confunde los errores tipo I y tipo II (distractor a) y el 6,5% de nuestros estudiantes confunde β y $1 - \beta$ (distractor b). El porcentaje en Vallecillos con este último error fue el 30%. Ella dio los distractores mediante símbolos; por ello puede ser más difícil que en nuestro caso, en que están dados en valor numérico.
- En el ítem 3, evalúa primordialmente si el alumno es capaz de diferenciar entre los errores tipo 1 y tipo 2 al plantear un test de hipótesis. El error tipo 2 se define como mantener una hipótesis nula que es falsa. En el caso que nos ocupa la hipótesis nula que se plantea es que las medias son iguales y se dice en el enunciado, que ésta hipótesis es falsa. De esta manera, si un investigador llega a la conclusión de mantener la hipótesis nula o que las medias no difieren, estará cometiendo un error tipo 2 (alternativa b, es la correcta). Por otro lado, si el investigador llegara a la conclusión de que las medias difieren y por tanto rechaza la hipótesis nula, no estará cometiendo ningún tipo de error, puesto que sabemos que la hipótesis nula es falsa (distractor c). En cuanto al distractor a, no es correcto, puesto que se habla de efectividad de los tratamientos, algo que no podemos inferir de las premisas que parte el enunciado, se puede decir que el alumno que elige esta alternativa está confundiendo hipótesis estadística e hipótesis de investigación (Chow, 1996). Otra explicación es que está suponiendo que un resultado estadísticamente significativo es ya un apoyo a la significación práctica (Pollard y Richardson, 1987; Abelson, 1997; Lecoutre, Lecoutre y Poitevineau, 2001, Díaz y Batanero, 2006).

Un 18,5% de nuestros alumnos elige el distractor c, (7,1% en Vallecillos). Hay un alto porcentaje y un 16,3%, de alumnos que no responden al ítem, lo que conlleva el desconocimiento entre Error Tipo I y Tipo II (20% en Vallecillos). De nuevo la autora da el enunciado en formato simbólico.

- En el ítem 4 se evalúa la comprensión del nivel de significación y el cálculo del valor crítico, así como la diferencia entre test unilateral y bilateral. La mayoría de los alumnos (46,7%) escoge el distractor b), lo que hacen es tomar el test unilateral izquierdo o lo que es lo mismo confunde hipótesis nula y alternativa o bien la región crítica y de aceptación en un contraste unilateral. Un 5,4% elige el distractor del mismo lado del contraste que se da en el enunciado, pero hace un error de lectura de tablas. Este ítem es de elaboración propia, aunque Tauber (2001) en su investigación sobre la distribución Normal muestra ítems de contenido similar, donde observa las dificultades que estudiantes de universitarios de primeros cursos tienen con el manejo de la distribución normal.
- En el ítem 5 evalúa el conocimiento del planteamiento de las hipótesis nula y alternativa, pero se centra más en la diferencia entre hipótesis estadística nula y alternativa y la propiedad que la hipótesis nula es siempre la de no diferencia. Un 3,3% de estudiantes de nuestro estudio no considera las hipótesis como complementarias. Vallecillos usa un ítem parecido, pero sólo presenta en cada distractor la hipótesis nula, por lo que no considera este tipo de error.
- En el ítem 6 evalúa el conocimiento de las condiciones en las que se tomará la decisión de rechazar la hipótesis nula. Un 2,2% elige el distractor b), correcto y un 26,1% el a), también correcto; pero no asocian que los dos enunciados responden a la misma pregunta. Un muy bajo porcentaje (13%) que no da una respuesta sobre la regla de decisión. Vallecillos usa un ítem parecido, pero plantea un ejemplo particular y no el caso general. Aún así sus resultados son peores.

4. Conclusiones

En cuanto al contenido referido al contraste de hipótesis nuestros resultados son bastante similares a los conseguidos por Vallecillos (1994) en su muestra de estudiantes de Psicología, incluso algo mejores. En líneas generales podemos considerar que los resultados han sido satisfactorios en la mayor parte de los ítems, con porcentajes de respuestas correctas bastante aceptables. Podemos afirmar que los alumnos son de asignar hipótesis estadística, dada una hipótesis de investigación y que éstos discriminan bien entre hipótesis nula y alternativa. La mayor parte de los participantes son capaces de hacer una asignación correcta de valor crítico y tienen un buen conocimiento de la regla de decisión.

Aun así, se conservan en nuestra muestra errores que tienen que ver con la confusión del nivel de significación, nivel de confianza, error tipo 1, error tipo 2 y potencia, y estas dificultades pueden ser tal vez explicado en parte a la al intercambio de sucesos que intervienen en la probabilidad condicional, tal y como apuntan Cohen, (1994); Falk y Greenbaum, (1995); Chow, (1996).

En este sentido apuntamos como idea la estrategia de incluir en el tema de inferencia estadística una clase en la que los alumnos exploren con mayor profundidad la idea de probabilidad condicional, pero desde el punto de vista “manipulativo” a través de applet disponibles en Internet. Por un lado, el uso de applet son atractivos para los alumnos, más que realizar problemas o cálculos y por otro lado permite que los alumnos experimenten diferentes posibilidades de forma más visual (Batanero, Contreras y Díaz, 2012).

Otra de las dificultades más relevantes encontradas tiene que ver con la elección del punto crítico para un nivel de significación especificado, confundiendo el procedimiento de hallar valores críticos en el contraste unilateral y bilateral, error no detectado por Vallecillos (1994). No queda claro si la confusión está entre contraste unilateral izquierda o derecha o tal vez en el manejo de las tablas de distribución normal tipificada. Sería interesante explorar mejor estas dificultades en investigaciones futuras.

Por otro lado, las variaciones que hemos introducido en los ítems confirman que estos errores no dependen del enunciado particular del ítem, sino de la falta de conocimiento de los estudiantes en estos conceptos. El enunciado verbal en varios de los ítems en lugar de simbólicos lleva a un mayor porcentaje confirmando la opinión de Falk (1986).

Referencias

- Alvarado, H. (2007). *Significados institucionales y personales del teorema central del límite en la enseñanza de estadística en ingeniería*. Tesis Doctoral. Universidad de Granada.
- Batanero, C. (2000). Controversies around significance tests. *Journal of Mathematics Thinking and Learning*, 2(1-2), 75-98.
- Batanero, C., Contreras, J. M. y Díaz, C. (2012). Sesgos en el razonamiento sobre probabilidad condicional e implicaciones para la enseñanza *Revista digital Matemática, Educación e Internet* 12(2).
- 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.
- Chow, L. S. (1996). *Statistical significance: Rationale, validity and utility*. London: Sage.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997 – 1003.
- Díaz, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el análisis de datos en psicología*. Tesis Doctoral. Universidad de Granada.

- Díaz, C. y de la Fuente, I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*, Volumen especial 2004, 161-167.
- Díaz, C. y Batanero, C. (2006). ¿Cómo puede el método bayesiano contribuir a la investigación en psicología y educación? *Paradigma*, 27(2), 35-53.
- Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning*, 9, 83 – 96.
- Falk, R. y Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology*, 5 (1), 75 – 98.
- Granaas, M. (2002). Hypothesis testing in psychology: throwing the baby out with the bathwater? En B. Phillips (Ed.). *Proceedings of the Sixth International Conference on Teaching of Statistics*. Ciudad del Cabo: International Association for Statistical Education. CD ROM.
- 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: Erlbaum.
- 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.
- Pollard, P., y Richardson, J. T. E. (1987). On the probability of making type I errors. *Psychological Bulletin*, 10, 159 – 163.
- Tauber, L. (2001). *Significado y comprensión de la distribución normal a partir de actividades de análisis de datos*. Tesis Doctoral. Universidad de Sevilla.
- Vallecillos, A (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral. Universidad de Granada.