

APRENDIZAJE Y ENSEÑANZA DEL CONTRASTE DE HIPÓTESIS: CONCEPCIONES Y ERRORES

VALLECILLOS JIMÉNEZ, A. y BATANERO BERNABEU, M.C.
Departamento de Didáctica de la Matemática. Universidad de Granada.

SUMMARY

In spite of the numerous references to problems derived from the incorrect use of the statistical tests or from misinterpretations of their results by experimental researchers in different areas, the educational world has stayed until the moment detached from them. In this paper we present the results of a wide research aiming at carrying out a comprehensive evaluation of the learning of the hypothesis test by undergraduates. We point to some difficulties and errors that might underlie the origin of the denounced problems and that the teaching will have to take into account for an improvement of the learning of the topic. Likewise, we describe students' conceptions about two key concepts in statistical test: the logical structure of hypothesis testing and the level of significance.

INTRODUCCIÓN

En el mundo de hoy la estadística está presente casi por doquier. Los medios de comunicación, los recursos informáticos disponibles y la formación básica general de los ciudadanos exigen conocimientos estadísticos de un cierto nivel que les permitan poder entender el mundo que nos rodea.

En el ámbito de la enseñanza, los profesores sienten desde hace tiempo la necesidad de la introducción de algunos conceptos estadísticos que permitan apreciar, interpretar y suministrar información de algunos fenómenos conocidos. Asimismo, son conscientes de la utilidad de las representaciones gráficas que pueden suministrar información sobre hechos o fenómenos de una manera concisa y adecuada. En el ámbito oficial, los nuevos planes de estudios para la enseñanza secundaria obligatoria y postobligatoria han incrementado los contenidos estadísticos, introduciendo temas novedosos como

el del contraste de hipótesis (Junta de Andalucía, 1992 y 1994). Ahora esto hace especialmente necesario que los profesores de estos niveles conozcan la nueva situación con el fin de poder afrontarla con renovada perspectiva (Vallecillos y Pérez-Ocón, 1995).

Por otra parte, recientemente, los medios informáticos han puesto al alcance de casi todo el mundo, no necesariamente con preparación específica en el tema, el uso de potentes recursos estadísticos de una manera fácil y rápida. Este hecho ha propiciado, sin duda, la aparición de algunos problemas derivados del uso de las técnicas estadísticas por incompreensión o abuso de las mismas (Morrison y Henkel, 1970; Batanero et al., 1992), pero también ha permitido, en el ámbito educativo, un mejor acceso al aprendizaje de algunos conceptos cuya comprensión es difícil, si no imposible, sin ellos (Vallecillos, 1995a).

En este trabajo describimos, muy sucintamente, los principales resultados obtenidos en una amplia investigación (Vallecillos, 1994) en la cual se han examinado de forma sistemática los errores conceptuales y procedimentales que se han producido en el aprendizaje del tema concreto del contraste de hipótesis por parte de una muestra grande de estudiantes universitarios. La descripción sistemática de los errores de los estudiantes en el aprendizaje de un tema (Batanero et al., 1994) es una valiosa información para los profesores y, mejor aún si se aportan algunas posibles causas o explicaciones para los mismos. Éstos tienen un poder de diagnóstico de los problemas (Borassi, 1987) con que se encuentran los estudiantes en el aprendizaje del tema y, por tanto, será conveniente que sean tenidos en cuenta en la planificación de una enseñanza que procure mejoras en la misma así como en sus aplicaciones. Pretendemos, por tanto, trasladar a los profesores algunos resultados de la investigación en el campo de la didáctica de la estadística, con el fin de contribuir a facilitar y mejorar su práctica docente. No tiene este trabajo, sin embargo, un carácter prescriptivo, sino orientativo para los profesores en el aula.

DESCRIPCIÓN DE LA INVESTIGACIÓN

La investigación de referencia (Vallecillos, 1994) ha tenido como fin hacer una valoración comprensiva y sistemática de las *concepciones* (Artigue, 1990), dificultades y errores de los estudiantes universitarios en el aprendizaje del tema del contraste de hipótesis, y se ha proyectado y llevado a cabo durante tres años, los cursos académicos del 1990-91 al 1992-93. Se ha planteado con carácter fundamentalmente descriptivo y orientativo y no pretende tener un valor absoluto extrapolable automáticamente a toda la población de estudiantes universitarios. Describimos a continuación sus aspectos generales más relevantes.

El tema que nos ocupa ha sido analizado, previamente a la realización de la valoración experimental sobre su aprendizaje, desde los diversos puntos de vista teóricos implicados: estadístico, filosófico-epistemológico e histórico (Vallecillos, 1994; Vallecillos, 1995b).

Desde el punto de vista estadístico, los profesores disponen de numerosos textos de referencia de los que citamos sólo algunos recientes como DeGroot (1988) o Canavos (1988). Sin embargo, para nuestro análisis del resultado del trabajo en el aula, esta información no es suficiente y, por tanto, hemos realizado un análisis conceptual didáctico basado, por una parte, en la teoría de los *actos de comprensión* de Sierpínska (1994) y, por otra, en la teoría del *significado de los objetos matemáticos* de Godino y Batanero (1994). Estos aspectos teóricos pueden ser consultados por los interesados en Vallecillos (1996b).

Las investigaciones previas a la nuestra han sido realizadas, principalmente, por investigadores preocupados

por el uso incorrecto de los tests de significación en su propio campo de trabajo (Morrison y Henkel, 1970) y no por especialistas del campo de la didáctica de las matemáticas. Algunos otros investigadores se han preocupado de las interpretaciones incorrectas del nivel de significación en los tests de hipótesis estadísticas en relación con la interpretación de un resultado significativo (Falk y Greenbaum, 1995) o en relación con la confusión e intercambio de los dos sucesos que intervienen en la definición del mismo, esto es, tomar como definición de α $P(H_0 \text{ sea cierta} \mid \text{rechazada } H_0) \neq \alpha$, en lugar de $P(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = \alpha$ (Birnbaum, 1982; Falk, 1986).

Objetivos e hipótesis

Sin embargo, en nuestra opinión, los errores e interpretaciones incorrectas del nivel de significación no pueden analizarse independientemente del resto de conceptos que intervienen en el contraste de hipótesis, por lo que es preciso realizar un estudio comprensivo y sistemático del aprendizaje del tema. En consecuencia nos marcamos como objetivo general la realización de un trabajo de estas características con una muestra representativa de estudiantes que nos permitiera estudiar de forma global y sistemática, por una parte, los tipos de errores relativos al contraste de hipótesis y su frecuencia entre los estudiantes universitarios de las distintas especialidades y, por otra parte, relacionarlos con las concepciones y errores relativos a otros conceptos implicados en el tema.

Las hipótesis básicas de nuestro trabajo son tres, sucintamente descritas a continuación (Vallecillos, 1994):

a) Las interpretaciones incorrectas del nivel de significación son diversas y no se reducen al intercambio de los sucesos que intervienen en su definición, descrito anteriormente por Birnbaum (1982) y Falk (1986).

b) Las interpretaciones incorrectas y errores referidos al nivel de significación dependen de errores referidos a otros conceptos clave que intervienen en el contraste de hipótesis, como, por ejemplo, el papel desempeñado por la distribución muestral del estadístico.

c) Los errores que afectan a conceptos varios y las interpretaciones incorrectas del nivel de significación descritas influyen decisivamente en los diversos errores procedimentales en la aplicación práctica de los contrastes.

Metodología

El trabajo experimental se desarrolló en tres fases netamente diferenciadas, estudio piloto, fase experimental y entrevista con un grupo de estudiantes escogidos, con el fin de profundizar en algunos aspectos y resultados anteriormente obtenidos y relevantes para la investigación (Vallecillos, 1994).

Para la toma de datos se ha empleado un cuestionario sobre dificultades conceptuales y procedimentales acerca del contraste de hipótesis, que no se incluye por razones de espacio, pero cuyo texto completo puede consultarse en Vallecillos (1996b). Consta de 20 ítems y un problema de aplicación de un contraste de hipótesis planteado en un contexto de control de calidad. En seis de los ítems se pide al estudiante que razone la respuesta dada al mismo con el fin de valorar mejor su comprensión del enunciado y la estabilidad de su respuesta. Los ítems se refieren a todos los conceptos clave que intervienen en el contraste de hipótesis, lógica y aplicación del proceso, establecimiento de hipótesis, fijación e interpretación del nivel de significación, errores de tipo I y II y su relación entre ellos, e interpretación de resultados. La prueba completa se ha construido con validez de contenido para evaluar el dominio investigado y comprobada su fiabilidad a través de los resultados obtenidos en una muestra piloto (Vallecillos, 1994).

A continuación se incluyen algunos ítems de la prueba, junto con un resumen del análisis de contenido previo realizado en cada uno de ellos. Con asterisco está señalada la respuesta correcta al ítem.

*Ítem 1: Un contraste estadístico de hipótesis, correctamente realizado, establece la verdad de una de las dos hipótesis nula o alternativa. V / *F*

Razona la respuesta.

Se trata de un ítem de verdadero/falso que se refiere a la lógica global del proceso para determinar si se diferencia entre la prueba de una hipótesis estadística y la demostración deductiva de una hipótesis lógica. Se pide además razonar la respuesta para comprobar la consistencia de la respuesta al ítem y poder profundizar en las concepciones subyacentes del alumno.

Ítem 2: ¿Cuál de las siguientes no es una hipótesis nula legítima?:

A: $\mu_x = 10$; B: $\sigma_x = 3$; *C: $\bar{x} = 35$; D: $\mu_1 = \mu_2$

Razona la respuesta.

Es un ítem de que se refiere al establecimiento de una hipótesis paramétrica. Los distractores A y B plantean la posibilidad de que la hipótesis se refiera a algún estadístico conocido (la media o la desviación típica) o a algún valor concreto de los mismos, ya que algunos estudiantes asocian el término «hipótesis nula» con el hecho de que el único valor admisible para la hipótesis sea cero, por ejemplo $\mu = 0$. El distractor D nos indicará los estudiantes que creen que una hipótesis nula debe ser expresada siempre mediante un único valor numérico; esto es, consideran sólo hipótesis nulas simples, o debe referirse a una sola población. Se pide razonar la respuesta.

*Ítem 19: Un nivel de significación del 5% significa que, en promedio, 5 de cada 100 veces que rechazamos la hipótesis nula estaremos equivocados. V / *F*

Razona la respuesta.

*Ítem 20: Un nivel de significación del 5% significa que, de promedio, 5 de cada 100 veces que la hipótesis nula es cierta, la rechazaremos. *V / F*

Razona la respuesta.

Estos dos ítems tienen una redacción sintácticamente muy parecida, si bien el primero es falso y el segundo verdadero, ya que se trata de la definición del nivel de significación. Se han planteado para estudiar, conjuntamente, la interpretación errónea del nivel de significación, que consiste en el intercambio de los dos sucesos, « H_0 cierta» y «rechazar H_0 » que intervienen en la definición del mismo; esto es, interpretar como a $P(H_0 \text{ sea cierta/rechazada } H_0) \neq \alpha$, en lugar de $P(\text{rechazar } H_0/H_0 \text{ cierta}) = \alpha$ (Birnbau, 1982; Falk, 1986). Pretendemos también caracterizar las concepciones de los estudiantes sobre el nivel de significación.

El cuestionario incluye además el siguiente problema.

Problema: La cantidad promedio que se coloca en un recipiente en un proceso de llenado se supone que es de 20 gramos. En forma periódica, se escogen al azar 25 recipientes y el contenido de cada uno de éstos se pesa. Se juzga el proceso como fuera de control cuando la media muestral \bar{x} es menor o igual a 19,8, o mayor o igual a 20,2 gramos. Se supone que la cantidad que se vacía en cada recipiente se encuentra aproximada, en forma adecuada, por una distribución normal con una desviación estándar de 0,5 gramos:

a) *Enunciar las hipótesis nula y alternativa que sean propias para esta situación.*

b) *Obtener la probabilidad del error de tipo I.*

c) *Calcular los nuevos valores críticos para x si aumentamos el tamaño de la muestra a $n = 36$, manteniendo el mismo nivel de significación.*

d) *Hacer una representación gráfica aproximada de la función de potencia para los siguientes valores medios de llenado: 19,6, 19,8 y 20.*

e) *Como una prueba alternativa considerar el rechazo de H_0 cuando $\bar{x} \leq 19,75$ o cuando $\bar{x} \geq 20,25$. Si el tamaño máximo aceptado para el error de tipo I es de 0,05, ¿cuál de las dos pruebas es la mejor si $\mu = 19,5$?*

f) *En los casos en que la hipótesis nula es rechazada, ¿existe la posibilidad de haber cometido un error? ¿Que tipo de error? ¿Cuál es la probabilidad de que estemos equivocados?*

El problema plantea la realización de un contraste de hipótesis en un contexto de control de calidad y se pide a los estudiantes establecer las hipótesis nula y alternativa adecuadas referidas a la media de la variable en la población de referencia. Una vez establecidas dichas hipótesis ($H_0: \mu = 20$; $H_1: \mu \neq 20$), el estudiante debe

calcular α utilizando su definición. Para ello debe determinar la media de la población, la distribución de la variable en la población y la distribución muestral de medias, $N(20, 0,5)$ y $N(20, 0,5/\sqrt{25})$, respectivamente. Posteriormente debe responder otras preguntas destinadas a comprobar el conocimiento de otros aspectos del contraste y el dominio del proceso al introducir algunas modificaciones en el mismo. El apartado *f* es de naturaleza distinta: se ha incluido en el problema con el fin de analizar, en otro contexto distinto, el error de intercambio de los condicionales en la definición del nivel de significación (Birnbaum, 1982; Falk, 1986), y completar los datos proporcionados por los ítems 19 y 20 del cuestionario.

El guión empleado como instrumento de recogida de datos en la entrevista está formado por tres ítems y una parte del problema de aplicación propuesto en el cuestionario anteriormente descrito. Se ha incluido además un problema de aplicación en un contexto de diagnóstico médico que puede ser interpretado en términos semejantes a un contraste de hipótesis, por lo que es especialmente indicado para valorar la comprensión y el grado de madurez conseguido por los estudiantes en el aprendizaje del tema. El texto íntegro del guión de entrevista no se incluye por razones de espacio pero puede consultarse igualmente en Vallecillos (1996b).

Descripción de la muestra experimental

El trabajo que describimos (Vallecillos, 1994) se ha llevado a cabo con la colaboración voluntaria de 436 estudiantes de la Universidad de Granada que han respondido el cuestionario y han resuelto el problema planteado. Pertenecen a catorce grupos de estudiantes y siete titulaciones diferentes: Ingenieros de Caminos, Canales y Puertos (93), Informática (63), Pedagogía (43), Psicología (70), Medicina (61), Matemáticas (31) y Empresariales (75). Son estudiantes de 2º, 4º y 5º curso de sus respectivas licenciaturas, esto es de primer y segundo ciclo; proceden de facultades con selección y sin ella para el acceso a sus estudios; los hay con diferente orientación en sus estudios, desde los más teóricos como los de Matemáticas, a los más aplicados como los de Psicología, pasando por otros más intermedios como Medicina y Empresariales. Todos ellos tienen incluidas en el currículo de sus estudios al menos una asignatura con contenido de probabilidades y estadística que incluye el tema del contraste de hipótesis. Su colaboración en la investigación ha sido voluntaria y desinteresada y conociendo los fines de la misma. Igualmente lo ha sido la de los profesores de los distintos grupos en cuyas clases se pasó el cuestionario. La muestra, aunque es intencional, la consideramos representativa de los estudiantes de la Universidad de Granada en ese momento, para los fines de la investigación.

La muestra empleada para la entrevista han sido siete estudiantes de medicina. Las razones para su selección han sido múltiples, aunque predomina la de su excelencia académica. Ya que los estudiantes seleccionados

deben ser capaces de explicar su trabajo y dado que el contenido es ciertamente complejo, necesitamos tener la seguridad de que los estudiantes seleccionados deben ser capaces de llevarlo a cabo en las mejores condiciones posibles. Por tanto deben ser estudiantes con rendimiento académico probado y además prepararse concienzudamente para la entrevista, por lo que se les advirtió de ella con antelación suficiente. Hay que señalar además que su colaboración fue excelente.

Una descripción detallada y los datos estadísticos de todas las muestras de estudiantes participantes, (distribución por grupos y especialidades, nivel de estudios y número de cursos de estadística previos, rendimiento en la asignatura de estadística en el curso actual e importancia y grado de dificultad subjetivamente atribuido al tema en estudio) pueden verse completos en Vallecillos (1994).

PRINCIPALES RESULTADOS OBTENIDOS

Dado el propósito de realizar un estudio comprensivo y sistemático del aprendizaje del tema objeto de la investigación, hemos estimado necesario realizar previamente un minucioso estudio teórico del tema y realizar una reorganización del mismo en una serie de aspectos o núcleos conceptuales que nos permitiera clasificar sistemáticamente las respuestas y los errores de los estudiantes. Los grandes núcleos conceptuales sobre los que hemos organizado nuestro análisis han sido los siguientes:

- a) El contraste de hipótesis como un problema de decisión: alternativas y consecuencias.
- b) Incertidumbre en los resultados: probabilidades de error y relación entre ellas.
- c) El nivel de significación y la potencia como riesgos del decisor.
- d) El parámetro y el estadístico de contraste.
- e) Nivel de significación y criterio de decisión.
- f) Nivel de significación y distribución del estadístico.
- g) Interpretación de resultados.
- h) Lógica global del proceso.

En cada uno de ellos se establecen, a su vez, una serie de aspectos distintos que conforman en total 32 aspectos diferentes a considerar en el contraste de hipótesis estadísticas. El cuestionario utilizado contiene ítems que se refieren a cada uno de los apartados anteriores. Sobre la base de este detallado análisis conceptual y didáctico del tema se establece la validez de contenido de la prueba realizada por los estudiantes. Para ello hemos realizado una tabulación cruzada de cada uno de estos aspectos y subspectos con cada uno de los ítems y el problema incluido en el cuestionario pasado a los estudiantes de manera que se ponga de manifiesto que están cubiertos con la prueba realizada todos los contenidos del tema que se pretenden analizar (Vallecillos, 1994).

Dificultades y errores encontrados en el aprendizaje del tema

En Vallecillos (1994) están recogidos con todo detalle los resultados teóricos obtenidos, así como los de los análisis estadísticos realizados, estudio de frecuencias por ítems, tabulaciones cruzadas por especialidades, análisis y clasificación de argumentos, análisis conjuntos de respuestas y argumentos, etc., que resumen los resultados completos, cuantitativos y cualitativos, obtenidos en la investigación realizada. Igualmente se recogen los resultados obtenidos en la fase de entrevistas. Describiremos aquí, a modo de síntesis de las conclusiones generales finales, muy resumidamente sólo aquellas dificultades y errores que han sido puestos de manifiesto por los estudiantes de una forma muy clara y generalizada. Los resultados se refieren a cada uno de los distintos núcleos conceptuales considerados en nuestro análisis conceptual del tema, resumido anteriormente, y se incluyen, además, algunos ejemplos de respuestas de los estudiantes que ilustren mejor el contenido de cada apartado.

En el apartado *a* se aprecia una gran dificultad de los estudiantes para establecer unas hipótesis adecuadas y para la diferenciación entre las hipótesis nula y alternativa que son frecuentemente intercambiadas (Vallecillos, en prensa). Estos errores conceptuales están en el origen de muchos de los errores producidos en la resolución del problema propuesto y creemos que, en general, pueden tener una gran influencia en la provocación de errores de tipo procedimental. En efecto, el establecimiento de las hipótesis adecuadas para una situación concreta en un paso inicial en un problema de contraste de hipótesis y el tomar como nula la hipótesis alternativa implica, en la práctica, tomar como nula una hipótesis compuesta, lo que bloqueará o impedirá a los estudiantes continuar con el problema. Por ejemplo, algunos estudiantes han dado como respuesta al apartado *a* del problema las siguientes hipótesis como adecuadas a la situación propuesta:

$$\begin{aligned} \text{«}H_0: \bar{x} = 20 \text{ gr o esta otra } \text{«}H_0: 19,8 \leq \bar{x} \leq 20,2 \\ H_1: \bar{x} \neq 20 \text{ gr} \text{»} \quad H_1: \bar{x} \leq 19,8 \text{ o } \bar{x} \geq 20,2 \text{»} \end{aligned}$$

En las dos respuestas anteriores se observan varias confusiones que se han repetido en numerosas ocasiones: en la primera se confunde el estadístico muestral \bar{x} con el parámetro poblacional μ ; en la segunda además se están tomando los datos del problema para construir la regla de decisión con objeto establecer unas supuestas hipótesis. El problema de la confusión de notación entre la empleada usualmente para referirse al estadístico muestral \bar{x} , y la empleada para el parámetro poblacional μ es generalizado. Al tratarse de una prueba escrita no podemos aclarar si se trata de una simple confusión en la notación o es la manifestación directamente observable de algún error conceptual. Por esta razón han sido interrogados al respecto los estudiantes entrevistados y nuestras conclusiones son que, al menos para estos alumnos, no se trata de una mera confusión de notación. En efecto, aunque los estudiantes distinguen perfectamente la me-

dia de la muestra y la de la población, manifiestan una falta de comprensión de los distintos papeles que juegan, en el contraste de hipótesis, el estadístico muestral y el parámetro poblacional. Se evidencia en todos los estudiantes entrevistados una falta de apreciación de la media muestral como variable aleatoria y de su distribución.

Por otra parte, en los resultados del cuestionario es manifestado asimismo, por numerosos estudiantes, una confusión entre los errores de tipo I y II que, por parte de algunos estudiantes, viene relacionada con el error anteriormente descrito de intercambiar el papel de las hipótesis, o en la relación entre los mismos, por ejemplo, se les considera complementarios o no incompatibles.

Respecto al apartado *b* encontramos fundamentalmente dos tipos de errores: las interpretaciones incorrectas del nivel de significación y los debidos a la falta de apreciación de la posibilidad de cometer un error en todo proceso de contraste. Entre estos últimos estarían los que consideran que el contraste finaliza con la prueba (lógica) de la hipótesis nula.

Los errores más notables que se refieren al apartado *c* son la consideración del nivel de significación como un valor fijo e inmutable y, por tanto, no calculable y las numerosas interpretaciones incorrectas del mismo. Como es sabido, el nivel de significación en un contraste de hipótesis se define como la probabilidad de rechazar la hipótesis nula siendo cierta. Hay pues, la probabilidad de cometer uno de los dos errores posibles, el llamado de tipo I y que simbólicamente escribimos así: $\alpha = P(\text{rechazar } H_0/H_0 \text{ cierta})$. El otro posible error sería la aceptación de una hipótesis nula falsa, que se denomina β y que simbólicamente escribimos así: $\beta = P(\text{aceptar } H_0/H_0 \text{ falsa})$. Se le llama error de tipo II. Hasta el momento, las referencias de investigaciones previas sobre el tema del contraste de hipótesis se han centrado en el concepto de *nivel de significación* y se han descrito los problemas derivados de su interpretación incorrecta como probabilidad de que la hipótesis nula sea cierta, una vez que ha sido rechazada, debido al intercambio de los sucesos que intervienen en su definición (Birnbaum, 1982; Falk, 1986). Para profundizar en el análisis de este y otros problemas, se han incluido en el cuestionario, entre otros, los ítems 19 y 20 citados con anterioridad. En el análisis de los argumentos que los estudiantes han dado como respuesta a estos ítems, nosotros hemos llegado a clasificar hasta quince interpretaciones distintas del nivel de significación (Vallecillos et al., 1992), y hemos analizado las posibles causas de algunos de los errores encontrados. En concreto, el error citado por Falk (1986), que consiste en interpretar como probabilidad de la hipótesis nula una vez rechazada, en nuestra opinión, parece estar relacionado más con la lógica global del proceso, por incompreensión de la misma, que con posibles dificultades de comprensión del concepto de *probabilidad condicional* (Vallecillos y Batanero, 1996b; Vallecillos y Batanero, en prensa). Por ejemplo, la siguiente respuesta textual de un estudiante expresa claramente

te la confusión del nivel de significación con la probabilidad de que ésta sea cierta una vez que ha sido rechazada:

– *No se trata de rechazar H_0 en un 5% sino de equivocarnos en un 5% cuando la rechazamos.*

En esta otra respuesta se manifiesta la confusión de α con la probabilidad de error en la decisión tomada (cualquiera que sea ésta):

– *Hay una probabilidad de acertar del 0,95. (Supone $\alpha = 5\%$).*

En el apartado *d* hemos encontrado numerosas respuestas erróneas de los estudiantes que evidencian una falta de apreciación de la variabilidad de la distribución del estadístico o, lo que es lo mismo, no perciben la media muestral como una variable aleatoria cuya distribución depende del parámetro poblacional. En muchos casos se están confundiendo el estadístico muestral con el parámetro poblacional (Vallecillos y Batanero, en prensa), como hemos comentado antes en relación con la situación en que se han producido estos errores.

En cuanto al nivel de significación como determinante de las regiones crítica y de aceptación y, por tanto, como criterio de decisión, (apartado *e*), y en relación con la distribución del estadístico (apartado *f*), hemos encontrado en muchas respuestas una confusión entre ambas regiones, y en relación con el tipo de contraste empleado, unilateral o bilateral. Análogamente los errores descritos en cuanto a la determinación e interpretación de las dos regiones, crítica y de aceptación, apuntan a una falta de comprensión del papel de la distribución del estadístico en el muestreo.

En el apartado siguiente (*g*), sobre la interpretación de resultados en un contraste, los errores más notables se refieren a la confusión de un resultado significativo, estadísticamente significativo o que lleva al rechazo de la hipótesis nula, con un resultado con relevancia práctica de algún tipo, lo que, evidentemente, nada tiene que ver con el resultado del contraste.

En el apartado sobre la lógica global del proceso de contraste (*h*), la confusión más notable es la que se refiere al tipo de validación de las hipótesis que proporciona el contraste que comentamos un poco más extensamente a continuación.

Concepciones sobre conceptos clave

Algunos de los ítems del cuestionario propuesto a los estudiantes, además de la elección de una opción como respuesta, incluyen la petición de razonar la respuesta dada, como hemos visto. Ésta es una pequeña parte de ensayo en donde el estudiante puede explicar muy brevemente, por razones del espacio que se les ofrece en el cuestionario o de tiempo entre otras, su respuesta. Estos argumentos proporcionados por los estudiantes nos han

permitido, tras un complejo y minucioso proceso de análisis, inferir las concepciones (Artigue, 1990) de los estudiantes sobre conceptos clave en el tema y tener acceso a una información sobre aspectos básicos generales del proceso que son muy importantes para la enseñanza del mismo. Una información más detallada de lo expuesto puede verse en Vallecillos (1995c; 1996a).

Hemos clasificado las concepciones referentes a dos aspectos clave del contraste de hipótesis: la lógica global del proceso y el nivel de significación.

1. Concepciones sobre el tipo de validación de las hipótesis que proporciona un contraste

• *Concepción del contraste como un criterio de decisión*

A modo de ejemplo del tipo de respuesta que soporta esta concepción, tendríamos las siguientes respuestas textuales de los estudiantes:

– *Lo que establece es la aceptación o el rechazo de una de las dos hipótesis.*

– *No es la verdad o no de una de las hipótesis lo que determina, sino que ofrece un criterio para aceptar o rechazar la hipótesis nula, dando además la probabilidad de equivocarnos.*

• *Concepción del contraste como mecanismo de apoyo empírico a una hipótesis bajo investigación*

Como ejemplo podemos citar las siguientes respuestas literales obtenidas:

– *El contraste de hipótesis se basa en inferir unas propiedades de la población a partir de datos muestrales. El resultado hace que se acepte una de las dos hipótesis, pero no quiere decir que sea verdad.*

– *El contraste de hipótesis no establece verdades. Se establece a través del contraste si una hipótesis, que se ha tomado como verdadera, es o no rechazable a través de los resultados obtenidos en un ensayo de población. (Destacado en el original).*

En los dos casos anteriores se trata de concepciones correctas, si bien en el primer caso se manifiesta un mayor acercamiento a la idea del contraste definido por Neyman-Pearson, mientras que el segundo podemos considerarlo más cercano a la filosofía de los tests de significación de Fisher (Lehmann, 1993).

• *Concepción del contraste como prueba probabilística de una hipótesis*

Esta concepción incorrecta del contraste de hipótesis manifiesta claramente la creencia en que el contraste proporcionará la «demostración probabilística» de la hipótesis, esto es, no la verdad sino la probabilidad *a posteriori* de la hipótesis con los datos disponibles. Obedece a un mecanismo de generalización abusiva (Artigue, 1990) del procedimiento de demostración

matemático que llamamos «por reducción al absurdo» o «por contradicción».

Las siguientes citas textuales de respuestas de estudiantes ilustran lo anterior:

– *Lo que se establece es la probabilidad, con un margen de error, de que una de las hipótesis sea cierta.*

– *Nunca se establece que sea verdadera una de las hipótesis sino que probablemente (con muchas posibilidades) sea cierta una de ellas.*

• *Concepción del contraste como demostración matemática de la hipótesis o prueba lógica de la misma.*

Esto conlleva el convencimiento de que, si se acepta una hipótesis, ésta es verdadera o, si se rechaza, es porque es falsa. Esta concepción incorrecta de la lógica del proceso es particularmente nefasta para la comprensión del mismo y para la correcta interpretación de los resultados del proceso en las aplicaciones. A modo de ejemplo del tipo de respuesta en la que subyace esta concepción tenemos las siguientes:

– *Porque antes de plantear el problema hay que establecer cuál es la hipótesis nula H_0 y la alternativa H_1 , y una de las dos ha de ser verdadera.*

– *Un contraste de hipótesis permite determinar si la hipótesis a contrastar (hipótesis nula) es falsa.*

2. Concepciones sobre el nivel de significación

El nivel de significación en un contraste de hipótesis es un concepto clave al que han dedicado tiempo y esfuerzo muchos investigadores preocupados por las interpretaciones incorrectas del mismo que ha llevado a graves errores de interpretación de los resultados de numerosas investigaciones, principalmente en el campo de la investigación experimental (Birnbaum, 1982; Falk, 1986; Morrison y Henkel, 1970; Vallecillos et al., 1992; Vallecillos y Batanero, 1995).

En nuestra investigación hemos encontrado muchas interpretaciones diferentes del mismo (Vallecillos et al. 1992) y las concepciones que describimos a continuación. Para ilustrar mejor el tipo de respuesta de los estudiantes que soportan cada una de ellas, se incluyen algunas respuestas textuales en cada caso.

• *Concepción correcta sobre el nivel de significación*

Hay muchos estudiantes que manifiestan correctamente la definición de *nivel de significación* como probabilidad de cometer el error de rechazar una hipótesis nula cierta, que contestan correctamente los ítems específicos y que escriben correctamente la probabilidad condicional correspondiente y la calculan, en su caso. Ejemplos:

– *5 de cada 100 veces que H_0 sea cierta, la rechazaremos.*

– *Porque $\alpha = P$ (rechazar H_0/H_0 cierta) pero no se supone que en esas 5 ocasiones que rechazamos H_0 , H_0 sea cierto.*

• *Concepción del nivel de significación como probabilidad condicional referida a alguna de las hipótesis*

Esta idea errónea se manifiesta, principalmente, en relación con la confusión entre los dos sucesos que intervienen en la definición, que son confundidos y se interpreta, en consecuencia, como definición de α , $P(H_0 \text{ sea cierta} / H_0 \text{ ha sido rechazada}) \neq \alpha$. Ejemplos:

– *Por definición de $\alpha = P(H_0 \text{ cierta/rechazarla})$.*

– *Es la probabilidad de equivocarnos al aceptar H_1 .*

• *Concepción del nivel de significación como probabilidad simple de la hipótesis nula.*

Algunos de los estudiantes interpretan el nivel de significación como probabilidad no condicional de la hipótesis nula, tanto para aceptarla como para rechazarla. Ejemplos:

– *El nivel de significación está o tiene que estar relacionado con el número de veces, o con el tanto por ciento que rechazamos la hipótesis nula.*

– *Por definición de nivel de significación P (rechazar H_0) $\leq 1 - \alpha$.*

• *Concepción del nivel de significación como probabilidad de error*

Hay numerosos estudiantes que conciben el nivel de significación como una simple probabilidad de error en la decisión tomada, tanto si se opta por la hipótesis nula o la alternativa como si se consideran ciertas o falsas. Ejemplos:

– *Lo que significa es que 5 de cada 100 veces nos equivocaremos en la hipótesis aceptada.*

– *Un nivel de significación del 0,05 significa que en el 95% tenemos la certeza de que no estamos equivocados.*

Finalmente, a modo de conclusión, haremos una breve reflexión sobre las implicaciones para la enseñanza del tema y de la estadística en distintos niveles educativos.

CONCLUSIONES

Los resultados obtenidos en nuestra investigación (Vallecillos, 1994), aunque muy sucintamente expuestos en este trabajo, evidencian la complejidad de un tema que no sólo es importante teóricamente sino de una gran importancia práctica como prueba su constante uso y abuso en las ciencias sociales y experimentales (Morrison y Henkel, 1970). Creemos que el hecho de que se

malinterpreten tanto los conceptos implicados como los propios resultados de la investigación basada en los tests de hipótesis apunta hacia la mejora en la enseñanza de los mismos como remedio de fondo para tratar de erradicar los problemas reseñados. Es necesaria, por tanto, una atención especial y abandonar la idea de la ilusión de transparencia de la comunicación didáctica, que hace que muchos profesores piensen que «mejorando la explicación» conseguirán «mejores resultados». Con este trabajo hemos conseguido señalar algunos auténticos «puntos negros» en el aprendizaje del tema, sobre los que es conveniente reflexionar para conseguir superarlos en la enseñanza.

Nos referiremos en primer lugar a algunos conceptos básicos que, a pesar de que se suponen comprendidos y aplicados, en su caso, en los niveles de enseñanza previos, es un hecho que se detectan fallos de comprensión que inciden directamente en la provocación de errores sobre el contraste de hipótesis. Principalmente nos referimos a los conceptos de *población* y *muestra* así como a la *variabilidad en el proceso de muestreo*; al concepto de *variable aleatoria* y su *distribución de probabilidad*. Es importante así mismo poner de manifiesto que en el proceso de contraste intervienen distintas medias (es el parámetro que interviene en nuestro caso): de la población, de la muestra y de la distribución muestral del estadístico. Para cada una de ellas se emplea una notación distinta que los estudiantes no llegan a discriminar y emplean a veces «al azar». Hemos observado que, en la prueba escrita, se produce una confusión generalizada, entre el parámetro poblacional y el estadístico muestral. Hemos podido comprobar posteriormente que, al menos con los estudiantes entrevistados, no se trata de una mera confusión entre las distintas notaciones sino que afecta a los conceptos y que en mayor medida apunta a la falta de comprensión del proceso de muestreo y a la distribución muestral del estadístico (la media en nuestro caso).

Creemos que es necesario también tener en cuenta que un número considerable de los estudiantes de nuestra muestra experimental, tras el proceso de enseñanza ordinario, ha llegado a tener unas concepciones incorrectas sobre el proceso del contraste de hipótesis que afectan a cuestiones fundamentales en estadística y en matemáticas. Por ejemplo, muchos de ellos piensan que se puede «demostrar» la hipótesis o «debilitan» un poco la conclusión y se quedan en «calcular la probabilidad de la hipótesis» mediante el proceso de contraste. La entrevista realizada en nuestro trabajo nos ha permitido profundizar en esta cuestión y buscar las posibles causas para

ello. Los resultados obtenidos nos han llevado a plantear la necesidad de realizar unas consideraciones epistemológicas previas sobre el papel de la inferencia estadística en la construcción del conocimiento (Rivadulla, 1991), que creemos necesario tener en cuenta para abordar con mayores garantías de éxito la enseñanza del tema y recomendar, por tanto, su introducción a los profesores encargados de llevarla a cabo (Vallecillos, 1995b).

Una parte importante de las respuestas de los estudiantes analizadas muestran interpretaciones incorrectas del nivel de significación que están directamente relacionadas con la incompreensión del tipo de validación de las hipótesis que proporciona el contraste. En nuestra opinión, este problema se produce por una falta de comprensión y de relación entre los conceptos de nivel de significación y regiones de aceptación y rechazo, enmarcado todo ello en la falta de apreciación de la distribución muestral del estadístico, que condiciona, a su vez, la comprensión de todo lo demás.

Otro aspecto a tener en cuenta es la manera en que los errores conceptuales afectan a las aplicaciones de los procedimientos de contraste. Una de las mayores dificultades encontradas ha sido el establecimiento de las hipótesis adecuadas a la situación planteada. Evidentemente ésta es una cuestión básica que condiciona fuertemente todo lo demás, por tanto, su incompreensión o errores relativos tienen una fuerte incidencia posterior, tanto para la enseñanza como para las aplicaciones de los contrastes de hipótesis por parte de usuarios de la estadística. Es necesario, pues, incidir en este punto en la enseñanza del tema, planteando problemas de tipo interpretativo, cercanos en lo posible a situaciones reales y que obligen a los estudiantes a reflexionar sobre las hipótesis necesarias al caso, frente a rutinarias aplicaciones del proceso o aplicación literal de fórmulas.

Por último, hacemos una llamada de atención en lo que se refiere al uso de la notación y las representaciones gráficas que, como lenguajes abstractos pueden representar dificultades añadidas para muchos estudiantes, especialmente principiantes.

AGRADECIMIENTOS

A todos los profesores y estudiantes que han colaborado desinteresadamente en pro de la enseñanza y de la investigación.

Este trabajo se realiza con financiación del Proyecto de Investigación PS93-0196 de la DGICYT (Madrid).

REFERENCIAS BIBLIOGRÁFICAS

- ARTIGUE, M. (1990). Epistemologie et Didactique. *Recherches en didactique des mathematiques*, 10(2-3), pp. 241-286.
- BATANERO, C., GODINO, J. y VALLECILLOS, A. (1992). El análisis de datos como útil y como objeto en didáctica de la matemática. *Educación Matemática*, 4(1), pp. 46-53.
- BATANERO, C., GODINO, J.; VALLECILLOS, A., GREEN, D. y HOLMES, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), pp. 527-547.
- BIRNBAUM, I. (1982). Interpreting statistical significance. *Teaching Statistics*, 4, pp. 24-27.
- BORASSI, R. (1987). Exploring mathematics through the analysis of errors. *For the Learning of Mathematics*, 7(3), pp. 2-8.
- CANAVOS, G.C. (1988). *Probabilidad y estadística. Aplicaciones y métodos*. México: McGraw-Hill.
- DEGROOT, M. H. (1988). *Probabilidad y Estadística*. México: Addison-Wesley Iberoamericana, SA.
- FALK, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning*, 9, pp. 83-96.
- FALK, R. y GREENBAUM, C.W. (1995). Significance tests die hard. *Theory & Psychology*, 5(1), pp. 75-98.
- GODINO, J. y BATANERO, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en didactique des mathematiques*, 14(3), pp. 325-355.
- JUNTA DE ANDALUCÍA (1992). Decreto 106/1992 de 9 de junio (BOJA del 20) por el que se establecen las enseñanzas correspondiente a la ESO en Andalucía.
- JUNTA DE ANDALUCÍA (1994). Decreto 126/1994 de 7 de junio (BOJA del 26 de julio) por el que se establecen las enseñanzas correspondientes al bachillerato en Andalucía.
- LEHMANN, E.L. (1993). The Neyman-Pearson theories of testing hypotheses: one theory or two? *Journal of the American Statistical Association*, 88, 424, pp. 1242-1249.
- MORRISON, D.E. y HENKEL, R.E. (eds.) (1970). *The Significance Tests Controversy - A Reader*. Chicago: Aldine.
- RIVADULLA, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropos.
- SIERPINSKA, A. (1994). *Understanding in mathematics*. Londres: The Falmer Press.
- 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. (1995a). Sugerencias metodológicas para la enseñanza del teorema central del límite en la enseñanza secundaria, en Ortega, C. J. A. (ed.). *El bachillerato, situación actual y perspectivas de futuro*, pp. 377-387. Granada: Fundación Educación y Futuro.
- VALLECILLOS, A. (1995b). Consideraciones epistemológicas sobre la inferencia estadística: implicaciones para la práctica docente. *Uno. Revista de Didáctica de las Matemáticas*, 5, pp. 80-90.
- VALLECILLOS, A. (1995c). Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios. *Recherches en Didactique des Mathematiques*, 15(3), pp. 53-81.
- VALLECILLOS, A. (1996a). Student's conceptions of the logic of hypothesis testing. *Hiroshima Journal of Mathematics Education*, 4, pp. 43-61.
- VALLECILLOS, A. (1996b). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis*. Granada: Comares.
- VALLECILLOS, A. (en prensa). El papel de las hipótesis estadísticas en los contrastes: concepciones y dificultades de aprendizaje. *Educación Matemática*.
- VALLECILLOS, A., BATANERO, C. y GODINO, J. (1992). Student's understanding of the significance level on statistical tests, en Geesling, W. y Graham, K. (eds.): *Proceedings of the XVI PME Conference*, 1, pp. 49-56. New Hampshire, Estados Unidos.
- VALLECILLOS, A. y BATANERO, M. C. (1995). La inferencia estadística en la investigación experimental en el campo educativo. *Revista de Educación de la Universidad de Granada*, 8, pp. 5-16.
- VALLECILLOS, A. y BATANERO, C. (1996). Conditional probability and the level of significance in the tests of hypotheses, en Puig, L. y Gutiérrez, A. (eds.). *Proceedings of the 20 PME*, 4, pp. 371-378. Valencia.
- VALLECILLOS, A. y BATANERO, C. (en prensa). Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios. *Recherches en Didactique des Mathematiques*.
- VALLECILLOS, A. y PÉREZ-OCÓN, R. (1995). Il curriculum statistico in Spagna: verso il 2000. La Statistica nell'istruzione secundaria. *Induzioni*, 10, pp. 43-52.

[Artículo recibido en marzo de 1996 y aceptado en abril de 1997.]

