

# CONCEPTOS ACTIVADOS EN EL CONTRASTE DE HIPÓTESIS ESTADÍSTICAS Y SU COMPRENSIÓN POR ESTUDIANTES UNIVERSITARIOS

*Angustias Vallecillos Jiménez y Carmen Batanero Bernabeu*

## RESUMEN

En este trabajo presentamos un estudio sobre las dificultades de comprensión de los estudiantes universitarios sobre algunos conceptos claves del contraste de hipótesis estadísticas: nivel de significación, hipótesis nula y alternativa, parámetro y estadístico, así como sobre la lógica global del contraste de hipótesis. Analizamos la comprensión de estos conceptos utilizando la técnica de entrevista llevada a cabo con siete universitarios que habían completado previamente un cuestionario escrito. Como consecuencia señalamos algunos puntos difíciles a los que sería necesario prestar especial atención en la enseñanza de los test de hipótesis.

## SUMMARY

In this paper we present a study of university students' difficulties in understanding key concepts of hypotheses testing. Using in-deep interviews of seven students who had previously completed a questionnaire, we analyze their understanding of the following concepts: significance level, null and alternative hypotheses, parameter and statistics, and the overall logic of hypotheses testing. As a consequence, we indicate some points needing special attention in the teaching of statistical tests.

## RÉSUMÉ

Dans cet article nous présentons une étude concernant les difficultés de compréhension d'étudiants universitaires sur quelques concepts clefs relatifs aux tests d'hypothèses. En menant un entretien clinique avec sept étudiants qui avaient rempli au préalable un questionnaire, nous avons analysé leur compréhension des concepts de niveau de signification, d'hypothèse nulle et alternative, de paramètre et statistique, ainsi que la logique globale du processus de test d'hypothèse. Nous en déduisons quelques points sur lesquels il serait nécessaire insister dans l'enseignement des tests d'hypothèses.

## 1. INTRODUCCIÓN

Los tests de hipótesis estadísticas han constituido una herramienta básica en la investigación experimental y, en particular, en la construcción de las ciencias de la conducta, no sin críticas respecto a su uso e interpretaciones incorrectas. Una síntesis de este debate se recoge en el libro editado por Morrison y Henkel (1970) y, más recientemente, en el artículo de Menon (1993) y las respuestas al mismo, o en Thompson (1996).

### *Conceptos implicados en un contraste de hipótesis estadísticas*

Para facilitar la comprensión de nuestro trabajo, incluimos a continuación un resumen de los principales conceptos que intervienen en el contraste de hipótesis, en la acepción de Neyman y Pearson, (vease, por ejemplo, en el libro de Zacks, 1981). Nos centraremos en el contraste de hipótesis paramétrica, ya que los conceptos que analizamos son comunes a otros contrastes, tales como los relacionados con la asociación entre variables o los de adherencia de una distribución empírica a un modelo teórico.

Un contraste de hipótesis es un procedimiento estadístico que permite decidir entre una de dos hipótesis complementarias  $H_0$  y  $H_1$  que, en el caso paramétrico, se refieren al valor constante, pero desconocido de un parámetro (e.g., la media o desviación estándar) de una variable aleatoria en una población. Para tomar la decisión, se obtienen un conjunto de valores de la variable  $(\mathbf{x}) = (x_1, x_2, \dots, x_n)$  en una muestra de individuos de esta población y, a partir de estos valores, se calcula el valor de un estadístico  $\tilde{v}(\mathbf{x})$ . Puesto que la muestra se elige aleatoriamente, el estadístico es una variable aleatoria, cuya distribución de probabilidad depende del valor del parámetro.

El conjunto de los valores posibles del estadístico se divide en dos regiones complementarias: la primera de estas regiones (región crítica  $C$ ) se elige de modo que la probabilidad de que el

estadístico tome valores en  $C$  es muy pequeña cuando la hipótesis nula  $H_0$  es cierta. La región complementaria  $A$  se denomina región de aceptación. Una vez tomada la muestra, si el valor particular observado del estadístico en ella pertenece a la región crítica, rechazaremos la hipótesis nula  $H_0$  y, en consecuencia, aceptaremos la alternativa  $H_1$ . El razonamiento seguido es el siguiente: o bien la hipótesis nula era cierta y ha ocurrido un valor de probabilidad muy baja, o bien la hipótesis nula era falsa.

La aplicación de un contraste de hipótesis puede dar origen a dos tipos diferentes de error: rechazar la una hipótesis nula que era cierta (error de tipo I) o aceptarla siendo falsa (error de tipo II). Aunque no podemos estar seguros de haber cometido uno de estos errores en un contraste particular, podemos determinar la probabilidad de rechazar la hipótesis nula, mediante la *función de potencia* del contraste  $P(\theta)$ , que se define en la forma siguiente, siendo  $\theta$  el valor desconocido del parámetro:

$$P(\theta) = P(\text{Rechazar } H_0 \mid \theta)$$

Si suponemos que nuestra hipótesis nula establece un valor dado para el parámetro  $\theta = \theta_0$  (caso de hipótesis nula simple), obtenemos el valor  $\alpha$ , nivel de significación del contraste o probabilidad de error de tipo I:

$$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ es cierta}) = P(\text{Rechazar } H_0 \mid \theta_0) = P(\theta_0)$$

También podemos calcular la probabilidad de aceptar  $H_0$  en función del parámetro:

$$\beta(\theta) = P(\text{Aceptar } H_0 \mid \theta)$$

En caso de aceptar la hipótesis nula, cometeremos un error de tipo II siempre que  $\theta$  sea diferente del valor supuesto  $\theta_0$ . Podemos ver que, en el caso de hipótesis nula simple, mientras que la probabilidad  $\alpha$  de cometer un error de tipo I es constante, la probabilidad  $\beta$  de cometer un error de tipo II depende del valor desconocido del parámetro. Finalmente, y puesto que los sucesos aceptar y rechazar la hipótesis nula son complementarios, se cumple la siguiente relación entre estas probabilidades:

$$\alpha = 1 - \beta(\theta_0)$$

Esta relación supone que al disminuir una de las probabilidades de error, la otra aumenta, si mantenemos un mismo tamaño de muestra.

#### *Antecedentes de nuestra investigación*

Muchos de los investigadores preocupados por el uso incorrecto de los contrastes de hipótesis en sus propias áreas de investigación sitúan el origen de los problemas denunciados en la interpretación incorrecta del concepto de nivel de significación (Bakan, 1966). La interpretación incorrecta más frecuentemente citada (Birnbaum, 1982; Falk, 1986; Vallecillos y cols. 1992) consiste en el intercambio entre los dos sucesos que intervienen en su definición, esto es, tomar como  $\alpha$  la probabilidad  $P(H_0 \text{ sea cierta} \mid H_0 \text{ ha sido rechazada})$ . Autores como Falk (1986) o Falk y Greenbaum (1995) han sugerido posibles causas para este error y razones para su persistencia en la investigación experimental, a pesar de los esfuerzos para su erradicación. Una posible causa es la creencia en que el nivel de significación se refiere a la probabilidad simple de un "*suceso condicionado*", auspiciada por la ambigüedad terminológica de la expresión "*probabilidad de error de tipo I*" usada al definir el nivel de significación. Otra posible explicación se refiere a la falta de diferenciación en ciertos sujetos entre la probabilidad condicional  $P(A \mid B)$  y  $P(B \mid A)$ , denominada "*falacia de la condicional traspuesta*" por Diaconis y Freedman (1982) o "*confusión de la inversa*" por Falk (1986). Ambas causas tienen su origen en una deficiente comprensión del concepto de probabilidad condicional. Otra razón de índole psicológica sugerida por Falk y Greenbaum (1995) y derivada de la necesidad de los sujetos de tomar decisiones en situaciones de incertidumbre, sería la creencia en que la obtención de un resultado significativo permite eludir el problema de la aleatoriedad en las mismas. Esta creencia se apoya en el

paralelismo (sólo aparente) de la lógica de los tests de hipótesis y la demostración por reducción al absurdo y en lo que ha sido denominada por Falk y Greenbaum la "*ilusión de la prueba probabilística por contradicción*", o creencia en que la obtención de un resultado significativo hace muy improbable la hipótesis nula.

## 2. OBJETIVOS Y METODOLOGÍA

Nuestra hipótesis particular es que los problemas denunciados de aplicaciones incorrectas o interpretaciones erróneas de los resultados de los contrastes de hipótesis, tienen también su raíz en una deficiente comprensión e interpretación de otros conceptos clave en el contraste de hipótesis, y no sólo del nivel de significación. Por otra parte, en lo que se refiere al nivel de significación, creemos que las interpretaciones incorrectas del mismo referenciadas no sólo son debidas a una deficiente comprensión del concepto de probabilidad condicional implicado en su definición sino también a otras causas diversas que iremos concretando en lo que sigue.

Con el fin de hacer una valoración experimental del aprendizaje del contraste de hipótesis hemos llevado a cabo una amplia investigación con estudiantes universitarios que se describe completa en Vallecillos (1994). En este trabajo presentaremos una parte de dicha investigación, que se refiere a la comprensión de la noción de probabilidad condicional, y de otros conceptos que intervienen en el contraste de hipótesis, así como de la lógica global de proceso de contraste, en relación con los errores relativos a la interpretación del nivel de significación.

En cuanto a la metodología seguida para llevar a cabo la investigación, ha sido una entrevista en profundidad con una muestra de siete estudiantes de 2º curso de Medicina seleccionados por su excelencia académica, y que han colaborado en forma ejemplar con la investigación. Previamente a la realización de la entrevista, se hizo una reunión con los alumnos participantes para informarles de los objetivos pretendidos con la misma y asegurarnos de que repasasen el tema, que habían estudiado con antelación. Las entrevistas fueron realizadas a cada uno de los alumnos individualmente por una de las investigadoras, inmediatamente después de que completaran un cuestionario escrito, pudiendo usar si lo deseaban calculadoras y tablas estadísticas. Las preguntas fueron hechas en tono coloquial y dejando hablar al entrevistado libremente. A continuación analizamos el cuestionario utilizado.

### *Cuestionario empleado en la entrevista*

En primer lugar el alumno debía indicar si cada uno de los siguientes enunciados era verdadero o falso, razonando la respuesta:

- Item 1:** Un contraste estadístico de hipótesis, correctamente aplicado, establece la verdad de una de las dos hipótesis nula o alternativa. (V/F\*).
- Item 2:** Un nivel de significación del 5% significa que, en promedio, 5 de cada 100 veces que rechazamos la hipótesis nula estaremos equivocados. (V/F\*).
- Item 3:** Un nivel de significación del 5 % significa que, en promedio, 5 de cada 100 veces que la hipótesis nula es cierta, la rechazaremos. (V\*/F).

Uno de los principales objetivos de la investigación ha sido realizar un estudio exploratorio cualitativo de las interpretaciones incorrectas que afectan al nivel de significación, en particular el error de intercambio de los dos sucesos que intervienen en su definición (Falk, 1986). Para ello hemos empleado los ítems 2 y 3, sintácticamente muy parecidos, que expresan esta confusión y han sido tomados de Birnbaum (1982). El ítem 3 es la expresión verbal de la definición del nivel de significación en un contraste de hipótesis, mientras que el ítem 2 invierte los sucesos en la probabilidad condicional que lo define. Birnbaum (1982) describe que muchos estudiantes toman el enunciado del ítem 2 como definición del nivel de significación al confundir las probabilidades presentadas en los ítems 2 y 3. En nuestra entrevista hemos preguntado a los estudiantes si creen que ambos enunciados son equivalentes y en caso de que no sea así, cuál de ellos se refiere a la definición del nivel de significación.

El ítem 1 fué usado en Vallecillos (1995) para analizar las concepciones de los estudiantes

sobre el tipo de prueba que proporciona el contraste de hipótesis y ha sido empleado para completar nuestro análisis de la comprensión del concepto por parte de estos estudiantes. Los estudiantes debían resolver, asimismo, dos problemas que analizamos a continuación.

**Problema 1:** La cantidad promedio de un cierto medicamento que se coloca en un envase en un proceso de llenado es de 20 gramos. Para el control de calidad, se eligen al azar 25 envases y el contenido de éstos se pesa. Se juzga que el proceso está 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) En el caso de haber rechazado la hipótesis nula,
  - c1) ¿existe la posibilidad de haber cometido un error?
  - c2) ¿que tipo de error?
  - c3) ¿cuál es la probabilidad de que estemos equivocados?

El enunciado presenta un problema típico de contraste estadístico sobre la media de una distribución normal, en un contexto de control de calidad. Se proporcionan como datos la media (20 grs.) y la desviación estándar (0.5 grs) del peso del producto que se desea embotellar en cada recipiente, en un proceso de llenado. El control de calidad se establece para comprobar el correcto llenado del recipiente y avisar de posibles desviaciones en el mismo. En consecuencia, si indicamos por  $\mu$  la media de la variable aleatoria considerada, que es un parámetro constante, pero desconocido, las hipótesis nula y alternativa adecuadas a la situación serían las siguientes:

$$H_0 : \mu = 20 \text{ grs.}$$

$$H_1 : \mu \neq 20 \text{ grs.}$$

El estadístico de contraste empleado en la situación descrita es la media  $\bar{x}$  del peso contenido en una muestra de 25 envases elegidos aleatoriamente. Esta media muestral es una variable aleatoria que tiene también una distribución normal, cuya media coincide con la de la variable en la población. En caso de que la hipótesis nula sea cierta (cuando  $\mu = 20$  grs.), la media de la distribución del estadístico sería 20 grs y su desviación estándar  $\sigma_{\bar{x}} = 0.5/\sqrt{25} = 0.1$ .

Puesto que el problema da como dato las regiones de aceptación ( $19.8 \leq \bar{x} \leq 20.2$ ) y crítica (la complementaria de la anterior), la probabilidad de error de tipo I se calcula fácilmente, tipificando la variable y usando las tablas de la distribución normal estándar o una calculadora científica que los alumnos manejan suficientemente:

$$\alpha = P(\text{Rechazar } H_0 \mid \mu = 20) = 1 - P(19.8 \leq \bar{x} \leq 20.2) = 1 - (P(-2 \leq Z \leq 2)) = 0.0456$$

Las tres últimas preguntas reiteran con diferente enunciado las preguntas contenidas en los ítems 1 y 2 y son de tipo interpretativo.

**Problema 2:** En un hospital se ha estudiado la efectividad de una prueba para detectar el cáncer con 3000 voluntarios, de los cuales 1000 son enfermos de cáncer y 2000 son personas sanas, obteniéndose los resultados que se muestran en la tabla 1.

Tabla 1: Resultado de la prueba diagnóstica

Tipo de sujeto	Positivo	Negativo	Total
Padece cáncer	900	100	1000
No padece cáncer	20	1980	2000
Total	920	2080	3000

- a) Si uno de estos sujetos padece cáncer, ¿cuál es la probabilidad de que la prueba resulte positiva?
- b) Si la prueba de uno de estos sujetos es positiva, ¿cuál es la probabilidad de que se trate de una persona que padece cáncer?
- c) ¿Qué tipos de errores de diagnóstico pueden cometerse con esta prueba? ¿Cuáles son sus probabilidades?

En este segundo problema presentamos a los alumnos, mediante una tabla de contingencia, datos relativos a la capacidad diagnóstica de una prueba de detección del cáncer. El contexto elegido, relacionado con la especialidad de los alumnos (estudiantes de medicina) permite dotar de significado práctico diferentes conceptos estadísticos, en particular los de probabilidad y probabilidad condicional, cuya comprensión deseamos evaluar en estos alumnos, mediante las preguntas a) y b). Finalmente, en el apartado c) se pide los errores posibles y sus probabilidades en un diagnóstico médico, que tiene una semejanza formal con un contraste de hipótesis, ya que se trataría de elegir entre las dos hipótesis siguientes:

$H_0$  = el sujeto está sano (hipótesis nula, de no efecto)

$H_1$  = el sujeto padece cáncer.

En este contexto, podemos comparar el error de Tipo I con la situación que se produce al dar un diagnóstico de cáncer a un sujeto sano (falso positivo) y el error de tipo II con la que ocurriría al dar un diagnóstico de sujeto sano a un enfermo de cáncer (falso negativo). Las probabilidades inversas (probabilidad de que habiendo tenido un resultado positivo en la prueba, se padezca cáncer o de que habiendo tenido un resultado negativo, se esté enfermo) pueden leerse directamente en la tabla, pues el problema se refiere sólo a los datos contenidos en ella. En caso de desear hacer una inferencia a la población serían preciso datos sobre la incidencia de la enfermedad y emplear la fórmula de Bayes.

La dificultad de este segundo problema respecto al primero es mucho menor, por el hecho de que tratamos con una variable dicotómica (tener o no tener cáncer) y no es preciso manejar los conceptos de regiones de aceptación o rechazo, estadístico y su distribución muestral.

### 3. ANÁLISIS DE LOS RESULTADOS OBTENIDOS

Las conclusiones obtenidas tras las entrevistas se han agrupado en diferentes categorías según los diferentes aspectos conceptuales o procedimentales del contraste de hipótesis, con objeto de identificar las concepciones de los alumnos (Artigue, 1990) ligadas a dichos objetos. A continuación presentamos nuestros resultados en cada una de estas categorías.

#### 3.1. INTERPRETACIÓN DE LAS PROBABILIDADES CONDICIONALES

Para estudiar la plausibilidad de las fuentes de error que afectan a la interpretación del nivel de significación citadas (Falk, 1986; Falk y Greenbaum, 1995) que señalan al concepto de probabilidad condicional como su principal origen, hemos interrogado al respecto a nuestros estudiantes para analizar el grado de comprensión manifestada por éstos sobre dicho concepto de probabilidad condicional.

Un primer punto a destacar es que estos alumnos no muestran dificultades en la resolución de los tres apartados del Problema 2, referido a la interpretación de probabilidades condicionales en una tabla de contingencia (Al. 1, 2, 3, 4, 5 y 6). Todos ellos reconocieron que la situación planteada en este problema requiere el uso de la probabilidad y fueron capaces de discriminar las probabilidades condicionales  $P(B|A)$  y  $P(A|B)$  en este contexto. Como ejemplo, el alumno 5 indica que "*son justamente la inversa, se diferencian fundamentalmente en el suceso condicionante, en el sentido de que, en la primera pregunta, el suceso condicionante es padecer cáncer y en la segunda, el suceso es que en la prueba,... el test dé positivo*".

En general, los alumnos fueron capaces de interpretar el Problema 2 desde la óptica del contraste de hipótesis. Incluso el alumno 3 cree que la semejanza entre ambos problemas se debe, fundamentalmente, a que existen dos tipos de errores y encuentra un paralelismo correcto de los falsos

positivos y falsos negativos con los errores de tipo I y II en un contraste de hipótesis. Aún más, el alumno 6 escribe, como solución del segundo apartado del problema 1, en donde se pide el cálculo de la probabilidad del error de tipo I, " $\alpha = P(H_1 | H_0) = P(H_1 \text{ y } H_0) / P(H_0)$ ". Es decir, no sólo es consciente de que  $\alpha$  viene dada por una probabilidad condicional, sino que escribe la fórmula general de cálculo de esta probabilidad.

Para el alumno 2 la diferencia entre una probabilidad simple y otra condicionada está en la dependencia de los sucesos que intervienen en la definición; si se tratase de sucesos independientes, la probabilidad condicionada se reduciría a una probabilidad simple. El alumno 6 aclara que, además, en una probabilidad condicional, la probabilidad depende de algo; sólo en el caso de sucesos independientes coincidirían la probabilidad simple y la condicional. El alumno 7 muestra un ejemplo concreto, aplicado a su especialidad "*por ejemplo una enfermedad y un factor de riesgo, para ver si ese factor de riesgo influencia o no a la enfermedad, eso sería la probabilidad condicionada y una probabilidad simple sería simplemente decir qué incidencia tiene una enfermedad en una población*".

Estos alumnos no consideran la existencia de un suceso condicionado; distinguen perfectamente que en la probabilidad condicional intervienen dos sucesos distintos. "*Un suceso condicionado?... no entiendo la pregunta... aquí hay un suceso condicionante y un suceso que está condicionado*", (Al. 5). Sólo el alumno 4 cree que puede haber casos en que en una probabilidad condicional intervenga un sólo suceso.

Por otra parte, tan sólo el alumno 7 muestra un error en la interpretación de las probabilidades condicionales en la tabla de contingencia que aparece en el enunciado del Problema 2, error identificado en la investigación de Estepa (1993) sobre la asociación estadística. Para contestar a la cuestión a), ¿cuál es la probabilidad de que la prueba resulte positiva?, aplica el teorema de Bayes, lo que puede entenderse como que ha interpretado la pregunta como relativa a una probabilidad condicional. Sin embargo, aunque emplea una notación correcta para la probabilidad pedida en el enunciado del problema  $P(C | A)$ , el valor que calcula es justamente el que se obtiene intercambiando los sucesos, esto es,  $P(A | C)$ .

En consecuencia, para los estudiantes de nuestra muestra, en general, no cabe admitir la ambigüedad verbal ni la insuficiente comprensión de la probabilidad condicional como causas potenciales de los errores de interpretación en el nivel de significación. Tan sólo en el caso del alumno 4 se aprecia una pequeña influencia de la ambigüedad verbal, en el sentido de creer que, en ocasiones, la probabilidad condicional podría ser identificada en base a un único suceso. Por otro lado, sólo el alumno 7 da muestras de intercambiar los condicionales en la definición del nivel de significación. Los otros 5 alumnos dan muestras de un alto grado de comprensión de la idea de probabilidad condicional, de discriminar las probabilidades condicionales, de aplicar el concepto en el Problema 2 y de buscar ejemplos significativos del mismo.

### 3.2. INTERPRETACIÓN DEL NIVEL DE SIGNIFICACIÓN

En general los estudiantes no encuentran equivalentes los enunciados de los ítems 2 y 3, aunque no porque tengan una concepción correcta del nivel de significación (Al. 1, 2, 3, 4, 5, 6). Podemos encontrar en estos alumnos una gran variedad de respuestas. El alumno 1 encuentra falsos los dos ítems. El alumno 5 encuentra que ambos son ciertos. El alumno 4 da respuestas correctas y consistentes a ambos ítems, diferenciando claramente los dos sucesos que intervienen en la definición de  $\alpha$ . Los alumnos 6 y 7 consideran que el ítem 2 es verdadero y el 3 falso, esto es, dan una respuesta errónea, pero consistente en ambos ítems, manifestando, la confusión de la inversa, descrita por Falk (1986). El alumno 6 aclara al ítem 3, "*5 de cada 100 veces que rechazamos  $H_0$  nos equivocaremos. Lo que se afirma en el enunciado es falso*". Este alumno percibe la diferencia entre la condición y el condicional en una probabilidad condicional y diferencia  $P(A | B)$  de  $P(B | A)$  en el caso general, aunque el alumno 7 confunde estas probabilidades condicionales.

Los alumnos 1, 2, 3, 5, 6 y 7 interpretan el nivel de significación  $\alpha$  como  $P(H_0 \text{ cierta} | H_0 \text{ ha sido rechazada})$ . Ello no se debe, no obstante, a una deficiente comprensión de la idea de probabilidad condicional, como sostiene Falk (1986). Todos los alumnos, excepto el 7, discriminan perfectamente las probabilidades  $P(A | B)$  y  $P(B | A)$  en el contexto del segundo problema y también diferencian

correctamente las ideas de probabilidad condicional y probabilidad simple. No hacen, pues, los estudiantes entrevistados, una identificación de las dos probabilidades condicionales  $P(\text{Rechazar } H_0 \mid H_0 \text{ cierta})$  y  $P(H_0 \text{ cierta} \mid H_0 \text{ ha sido rechazada})$ . Por el contrario, estos alumnos interpretan el nivel de significación como la probabilidad a posteriori de la hipótesis nula, en función de los datos obtenidos.

El nivel de significación  $\alpha$  es, para todos los alumnos entrevistados, una probabilidad condicional que se interpreta como probabilidad de error en caso de rechazar la hipótesis nula, esto es, se invierten los términos en la probabilidad condicional que define  $\alpha$ . (Al. 1, 6 en las respuestas a los ítems 2 y 3). Asimismo esta creencia se pone de manifiesto en los alumnos 1, 2, 3 y 5 al contestar al apartado c) del Problema 1. Como veremos a lo largo del estudio, las causas que provocan este error son principalmente que los alumnos creen que el contraste estadístico es un procedimiento que permite calcular inductivamente la probabilidad a posteriori de la hipótesis y errores conceptuales relacionados con el estadístico de contraste y su distribución.

Aún más, los alumnos 1, 2, 3, 4, 5 y 6 son capaces de interpretar correctamente las definiciones de  $\alpha$  y  $\beta$  en el contexto del segundo problema, aunque, como el caso de los alumnos 2 y 6, en un primer intento, no encuentran semejanza entre los problemas 1 y 2. Sin embargo, no son capaces de transferir esta interpretación correcta al problema del contraste de hipótesis ni a los ítems que se refieren a él.

En algunos casos, como los alumnos 2 y 7, se manifiesta la confusión de los condicionales en la definición de  $\alpha$  debida a la ambigüedad verbal y no a la falta de comprensión de la idea de probabilidad condicional. "*Si se acepta  $H_1$ , entonces el error que puedes cometer ... es el error  $\alpha$* " (Al. 7). Creemos que este es también el caso del alumno 4, que da una definición correcta de  $\alpha$ , pero, finalmente, lo interpreta incorrectamente en el apartado c) del Problema 1.

Un matiz importante, manifestado por el alumno 3, es que algunos alumnos encuentran falso el ítem 3, porque para ellos el error tipo I y su probabilidad  $\alpha$  deben estar asociado a una decisión relacionada con  $H_0$  y las decisiones relacionadas con  $H_1$  se referirían al error tipo II y su probabilidad  $\beta$ . Se muestra una confusión entre los errores tipo I y II debido a la falta de comprensión de la equivalencia de "*aceptar  $H_1$* " y "*rechazar  $H_0$* ". Esto se explica, en el caso concreto de los alumnos entrevistados, porque en la enseñanza recibida normalmente han usado la terminología "*aceptar  $H_1$* ", en lugar de "*rechazar  $H_0$* " y este hecho resulta relevante para la correcta comprensión del concepto.

### *Cálculo del nivel de significación*

En general, los alumnos no llegan a completar correctamente el cálculo del nivel de significación en el Problema 1. El alumno 1, aunque realiza el cálculo, cree que se refiere a la probabilidad de la hipótesis nula, en consistencia con su creencia de que este es el objetivo final del contraste de hipótesis y con la identificación que hace de la hipótesis nula y alternativa, así como de las regiones de aceptación y rechazo. En el caso del alumno 7, él cree que el nivel de significación no se puede calcular, que es un valor elegido por el investigador antes de realizar el contraste, pero no calculable a partir del conocimiento de las regiones críticas y de aceptación.

Los que han efectuado el cálculo de  $\alpha$  utilizando la definición del mismo, (Al. 2 y 3), han cometido todos ellos un mismo error de tipificación: dividir por 0'5 en lugar de por 0'5/5, esto es, no han tenido en cuenta para nada la distribución del estadístico ni el tamaño de la muestra, sino que han usado la distribución de la variable en la población. A pesar de que todos ellos saben que  $\alpha$  es un valor pequeño y tienen la referencia concreta de los valores 0'05 y 0'01, no ha llamado su atención el alto valor obtenido ( $\alpha=0.6892$ ). Estos resultados parecen indicar que los alumnos no han conseguido poner en relación el nivel de significación con la distribución del estadístico.

El resto de los estudiantes (Al. 4, 5, y 6) se limitan a representar simbólicamente la definición de  $\alpha$ , sin calcularlo, reconociendo, en algunos casos que no saben efectuar éste cálculo. Concluimos, en consecuencia, que son dos las principales causas que hacen que los alumnos de esta muestra interpreten incorrectamente el nivel de significación:

a) La creencia en que el contraste de hipótesis es un procedimiento que permite el cálculo inductivo de la probabilidad a posteriori de la hipótesis, en función de los datos. Esto es, los

estudiantes manifiestan una *concepción del contraste de hipótesis como prueba probabilística de la hipótesis* (Vallecillos, 1995), por realizar una generalización abusiva de los conocimientos que poseen sobre la demostración por reducción al absurdo (Artigue, 1990).

b) La falta de apreciación de la idea de estadístico como variable aleatoria, de la dependencia de su distribución, en función de la hipótesis y de la relación de esta distribución con el nivel de significación. Posiblemente hay aquí un mecanismo de aplicación de la heurística de la representatividad, (Kahneman y cols., 1982) y el alumno identifica los valores muestrales obtenidos con los de la población desconocida; juzga así, la probabilidad del parámetro desconocido (al cual se refiere la hipótesis) en función de los datos obtenidos, que espera sean similares al valor de dicho parámetro.

No obstante, esporádicamente, algunos alumnos, como el 7, han mostrado una confusión entre las dos probabilidades condicionales implicadas, lo que puede influenciar también su interpretación incorrecta del nivel de significación. Por tanto, tampoco podemos descartar completamente esta última causa como posible inductora de algunos de los errores que ocasionan interpretaciones incorrectas del nivel de significación.

### 3.3. CONCEPTO DE HIPÓTESIS: HIPÓTESIS NULA Y ALTERNATIVA

Puesto que en nuestros alumnos no puede sostenerse la hipótesis de que las dificultades de interpretación del nivel de significación sean debidas a una interpretación inadecuada de la probabilidad condicional, hemos explorado su comprensión de otros conceptos relacionados con el nivel de significación. En primer lugar, analizamos su comprensión de la idea de hipótesis y sus tipos, ya que existen diferencias fundamentales entre el concepto de hipótesis estadística, respecto al sentido dado al término hipótesis, en el contexto de la investigación (Scott, 1988), aspecto que no siempre es suficientemente comprendido por los alumnos.

Algunos de los alumnos entrevistados, consideran las hipótesis como alternativas en un problema de decisión bajo incertidumbre, de las cuales no se conoce su certeza (Alumno 1). Sin embargo, la mayoría no establece el paralelismo entre el contraste de hipótesis y el proceso de decisión. Para ellos la hipótesis es un enunciado o situación desconocida de la que se desea determinar su verdad o falsedad (Al. 4, 7). También lo expresan diciendo que la hipótesis es un enunciado a probar o rebatir, una afirmación sujeta a demostración (Al. 2, 3, 5, 6). Esta concepción del papel que juegan las hipótesis estadísticas se aparta de la lógica del contraste de hipótesis que usualmente se trata en la enseñanza de estos alumnos como proceso de decisión para aceptar o rechazar una hipótesis, acorde con la teoría de Neyman y Pearson.

Un error detectado en algunos de estos alumnos es la identificación que hacen de la hipótesis nula con la región de aceptación y de la hipótesis alternativa con la región de rechazo. Puesto que una hipótesis se acepta o rechaza en función de que el resultado del contraste pertenezca a la región crítica o de aceptación, algunos alumnos identifican estos conceptos, (Al. 1, 3, 5). Este error se pone de manifiesto en la práctica por la elección de hipótesis que realizan en el apartado a) del Problema 1. Incluso, algún alumno, como el 5, declara expresamente que una hipótesis y su región de aceptación *"son la misma cosa"*.

Otro error conceptual identificado en las respuestas de los estudiantes es la creencia en que una hipótesis puede referirse indistintamente a una población o a una muestra (Hawkins, y cols., 1992). (Al. 1, 2, 3). Ello hace que, en el Problema 1, los alumnos utilicen preferentemente la notación  $x$  en lugar de  $\mu$  al plantear las hipótesis requeridas (Al. 6). Sin embargo algunos de estos alumnos distinguen entre dato e hipótesis; la hipótesis es un valor supuesto, aunque se refiera a la muestra; el dato es un valor conocido, incluso cuando se refiera a la población.

Como veremos más adelante, es el hecho de que el alumno no considera la media muestral como una variable aleatoria, lo que puede ocasionar la creencia en que la hipótesis pueda referirse indistintamente a la media de la muestra, tanto como a la de la población. Es decir, aunque los alumnos discriminan ambas medias (Schuyten, 1991), no son conscientes de que la media de la población es una constante, mientras que la media de la muestra es una variable aleatoria.

### *Elección de las hipótesis en los problemas propuestos*

Los alumnos 4 y 7 plantean correctamente las hipótesis nula y alternativa en el apartado a) del Problema 1. El segundo, incluso, explica con claridad su elección, dando muestras también de diferenciar entre el test bilateral y unilateral. Los demás, en general, no plantean correctamente las hipótesis adecuadas en los problemas propuestos por problemas diversos.

Los alumnos 1 y 3 toman como hipótesis las regiones de aceptación y rechazo dadas en el enunciado como habíamos indicado con anterioridad. Los alumnos 2 y 6 establecen correctamente el valor hipotético supuesto en el enunciado del primer problema, pero creen que este valor hipotético se refiere a la media de la muestra y no a la media de la población. También el alumno 5 elige sus hipótesis haciendo referencia a la media de la muestra y no a la media de la población.

El alumno 5 toma unas hipótesis nula y alternativa que no completan el espacio paramétrico y, aparentemente, la hipótesis nula está incluida en la alternativa; esto es, se trata de hipótesis no excluyentes ni complementarias. Muestra también confusión con las regiones de aceptación y rechazo.

Otro problema detectado en estos alumnos es que, aunque a nivel teórico declaran que la hipótesis nula se plantea para ser rechazada, no son consecuentes cuando, a lo largo de la entrevista, se les pide plantear la hipótesis nula en el Problema 2. El contexto del problema indica que se dispone de una prueba de diagnóstico del cáncer; esta prueba se aplicará, usualmente, cuando haya sospechas de que una persona tenga esta enfermedad. Esto es, se investiga la hipótesis de que el sujeto padezca cáncer usando una determinada prueba diagnóstica. Puesto que lo que deseamos es confirmar nuestras sospechas, se debiera tomar como hipótesis nula en un supuesto contraste el "*no tener cáncer*", es decir, la contraria a la hipótesis bajo investigación. Si, bajo esta hipótesis, obtuviésemos un resultado muy improbable, se rechazaría la hipótesis bajo investigación, o, lo que es lo mismo, se daría un diagnóstico de cáncer.

Hay otro hecho a favor de tomar como hipótesis nula la de "no efecto" (en este caso no tener cáncer). Como sabemos, en la lógica del contraste, los cálculos se realizan bajo la suposición de que la hipótesis nula sea cierta. Todos estos cálculos serían inválidos si tomásemos como hipótesis nula la más improbable. Ahora bien, sabemos que de las dos hipótesis, tener o no tener cáncer, la primera es más improbable, utilizando una definición frecuencial de la probabilidad.

Sin embargo, en el contexto del Problema 2 y, cuando se pide a los alumnos que identifiquen los errores de tipo I y II y sus probabilidades en el mismo, una parte de ellos han tomado como hipótesis nula el "*tener cáncer*" (Al. 1, 5, 6). En el caso de los alumnos 2, 3 y 4 toman como hipótesis nula del segundo problema el "*no tener cáncer*", dando, en consecuencia, muestras de una capacidad de síntesis de los diversos conceptos a que hemos hecho mención y de transferencia de los mismos a este nuevo contexto. Calculan también correctamente  $\alpha$  y  $\beta$ , por lo que podemos considerarlos con una alta capacidad procedimental. El alumno 7, con una gran dificultad, identifica finalmente  $\alpha$  y  $\beta$  en el contexto de este problema.

En resumen, los alumnos consideran las hipótesis estadísticas como enunciados sujetos a demostración, discriminando entre dato e hipótesis, así como entre hipótesis nula y alternativa. Algunos de ellos creen que una hipótesis puede referirse indistintamente a la muestra o a la población, aunque discriminan estos dos conceptos, o hacen una limitación excesiva de las hipótesis nulas admisibles. En general, observamos una dificultad en los alumnos entrevistados para elegir las hipótesis adecuadas, a partir del enunciado verbal de los problemas motivado por la adquisición insuficiente de algunos elementos de significado (Godino y Batanero, 1994) relacionados con el concepto de hipótesis estadística. Llamamos la atención sobre este hecho, puesto que la elección de unas hipótesis inadecuadas condiciona los resultados del resto del procedimiento. Por otra parte, y referido a las aplicaciones, es una de las causas de la dificultad en la comunicación de estadísticos y usuarios de la Estadística.

### 3.4. LÓGICA GLOBAL DEL PROCESO

Otro punto analizado fueron las concepciones de los alumnos de nuestra muestra sobre la lógica global del proceso de contraste, esto es, sobre el tipo de validación de las hipótesis que éste

proporciona. Haremos solamente un resumen, puesto que los resultados han sido publicados con anterioridad en un trabajo que trata específicamente de las concepciones de los alumnos respecto a la lógica global de un contraste de hipótesis y su relación con el problema de la inducción en las ciencias empíricas (Vallecillos, 1995).

Una concepción que se destaca sobre el tipo de prueba que proporciona un contraste de hipótesis es la de que éste procedimiento, aunque no prueba la hipótesis, permite calcular un grado de certeza o probabilidad de la misma. Es manifestada con diversas variantes por todos los alumnos: En el caso del alumno 1, esta creencia se refuerza por la interpretación que da a las probabilidades asociadas a las regiones de aceptación y crítica, como probabilidades, respectivamente, de certeza de las dos hipótesis nula y alternativa. En algunos casos, como los de los estudiantes 5, 6 y 7, comparan el contraste de hipótesis con la demostración de un teorema matemático, aunque, sugieren, su demostración está sometida a las "*probabilidades, a los errores*".

Los alumnos creen, en consecuencia que como resultado final del procedimiento de contraste, se obtendrá la probabilidad de la hipótesis. El cálculo de esta probabilidad se obtendría aplicando los cálculos que realizan durante el contraste de hipótesis; para ellos, en consecuencia, la probabilidad a posteriori de la hipótesis sería la probabilidad del resultado obtenido, bajo la condición de la hipótesis nula cierta, esto es, lo que conocemos como p-valor. Una acepción diferente dada a la probabilidad de una hipótesis es la del alumno 3 quien sugiere como método de cálculo de la probabilidad de la hipótesis tomar varias muestras y estudiar la frecuencia en las que se verifica la hipótesis. Esto es, para él la probabilidad de la hipótesis sería estimable mediante una aproximación frecuencial.

En resumen, los alumnos entrevistados, en mayor o menor grado, comparten una *concepción del contraste de hipótesis como prueba probabilística de la hipótesis* (Vallecillos, 1995), aunque con diversos matices. Mientras que el alumno 3 cree que se llegará a obtener la probabilidad de la hipótesis, a partir de un método frecuencial, mediante experimentación repetida, el resto de los alumnos consideran que son los cálculos de probabilidad realizados durante el procedimiento de contraste, los que proporcionan la probabilidad de la hipótesis, entendida en su acepción lógica.

### 3.5. PARÁMETRO, ESTADÍSTICO DE CONTRASTE Y SU DISTRIBUCIÓN

Un error cometido por casi todos los estudiantes, ha sido la confusión entre estadístico muestral y parámetro poblacional. En general los alumnos diferencian teóricamente la media particular de la muestra dada de la media de la población (Al. 1, 2), aunque no identifican en ninguno de los casos otras medias implicadas en el procedimiento de contraste.

El alumno 1 diferencia los distintos tipos de medias, aunque no encuentra necesidad de emplearlas en el problema; posiblemente debido a su falta de comprensión del papel de la distribución del estadístico en el proceso. Para él los cálculos que se realizan se refieren a la probabilidad de la hipótesis y no a la probabilidad de obtener un valor del estadístico, supuesta cierta dicha hipótesis.

Ante las preguntas de la investigadora sobre otras posibles medias intervinientes en el proceso de contraste, el alumno 2 sólo es capaz de pensar que se está refiriendo a la media de la distribución normal  $N(0, 1)$ , que debe utilizar como instrumento, al tener que tomar las tablas de ésta distribución para calcular las probabilidades que intervienen en el problema. El alumno 3 encuentra suficiente la media muestral para resolver las diversas partes del problema. Aunque diferencia entre media de la muestra y de la población, no diferencia el papel jugado por cada una de ellas en el procedimiento de contraste. Lo mismo ocurre con los alumnos 4, 5 y 6 que formulan las hipótesis empleando la media muestral, en lugar de la poblacional.

Como hemos podido observar en el apartado referido a las concepciones de los estudiantes sobre las hipótesis estadísticas, los alumnos creen que éstas pueden referirse indistintamente al estadístico o al parámetro y ello hace que utilicen, preferentemente, la notación de la media muestral en lugar de la poblacional.

Estas respuestas nos confirman en que la confusión detectada entre estadístico muestral y parámetro poblacional no es una confusión de notación únicamente, sino que obedece a una falta de discriminación real entre ambos conceptos. Es decir, los alumnos no identifican la variable aleatoria "media muestral" que se obtendría al variar las muestras en la población. No son conscientes tampoco

del papel jugado por la media de esta variable aleatoria, esto es, por la esperanza matemática de la distribución muestral del estadístico. Hay también una falta de comprensión del estadístico como variable aleatoria y de la función que desempeña la distribución del estadístico en la determinación de las regiones crítica y de aceptación.

Los alumnos 1, 2 y 3 no utilizan ni hacen referencia a esta distribución muestral. Por ello, en los cálculos realizados emplean, en lugar de esta distribución, la distribución de la variable dada en la población. Ello hace que obtengan unos valores anormalmente grandes de la probabilidad. El resto de los alumnos no llega a calcular  $\alpha$ .

Asimismo hay una deficiente comprensión de la relación entre la distribución del estadístico, como función de la hipótesis nula y la posibilidad de rechazar erróneamente un porcentaje de hipótesis nulas ciertas (Al. 1). Igual ocurre con el alumno 3, que no entiende como es posible, que entre 100 hipótesis nulas ciertas se rechacen 5. También el alumno 7 interpreta erróneamente el enunciado del ítem 3 debido a la incomprensión del papel de la distribución del estadístico en la probabilidad de error. Para él  $\alpha$  es "*un error y no un porcentaje de rechazos*".

Además, en algunos casos, como el alumno 3, calculan correctamente el nivel de significación  $\alpha$ , salvo que usan la distribución de la población en lugar de la del estadístico. Aunque el cálculo que hacen pudiera dar a entender que comprenden la dependencia de la distribución del estadístico, respecto al valor supuesto del parámetro, esta comprensión es, a nuestro juicio, aparente, puesto que no usan en el cálculo del nivel de significación la hipótesis establecida en el primer apartado del problema.

Deducimos, en consecuencia, que el alumno ha sido capaz de memorizar el algoritmo de cálculo, e incluso de transferirlo al problema, eligiendo los datos pertinentes. No ha sido, sin embargo, capaz de lograr una comprensión conceptual profunda de los teoremas y conceptos en acto (Vergnaud, 1982; 1990) subyacentes en el procedimiento de cálculo.

#### 4. CONCLUSIONES

En los apartados anteriores hemos puesto de manifiesto, mediante el análisis de entrevistas en profundidad llevadas a cabo con una muestra de siete estudiantes universitarios, la existencia de errores en la interpretación del nivel de significación en alumnos con una buena preparación estadística y comprensión de la idea de probabilidad condicional. Dos son las principales causas que hacen que los alumnos de esta muestra interpreten incorrectamente el nivel de significación:

a) La creencia en que el contraste de hipótesis es un procedimiento estadístico que permite el cálculo inductivo de la probabilidad a posteriori de la hipótesis, en función de los datos, y

b) La falta de apreciación de la característica de variable aleatoria del estadístico muestral. Todos los alumnos de la muestra ha cometido errores que evidencian falta de comprensión referidas a las relaciones entre la distribución del estadístico, las regiones y el nivel de significación. Ninguno de ellos parece comprender correctamente el carácter de variable aleatoria del estadístico, lo que, en parte, explica también la interpretación que anteriormente han dado a los cálculos que se efectúan durante el proceso de contraste, como cálculos conducentes a la obtención de la probabilidad (a posteriori) de la hipótesis.

Estos resultados no contradicen los descritos por Falk y Greenbaum (1995), si bien las causas que originan los problemas descritos, en nuestra opinión, no coinciden, al menos totalmente, con las supuestas por estos autores. Aunque es posible la existencia de los mecanismos psicológicos que condicionan el comportamiento de los sujetos en situaciones de incertidumbre, la causa fundamental, en nuestra opinión, es la falta de consecución de los actos de comprensión que hemos señalado, así como otros errores sobre el concepto de hipótesis, lo que contribuye a la permanencia de la "*ilusión de la prueba probabilística por contradicción*" (Falk y Greenbaum, 1995) y demás errores de aplicación de los tests descritos por estos autores. A falta de éstos elementos de significado, el significado personal del contraste de hipótesis para el alumno no es coincidente con el significado en la institución matemática (Godino y Batanero, 1994).

Posiblemente a ello se añada un mecanismo de aplicación de la heurística de la representatividad, (Kahneman y cols., 1982) y el alumno identifica los valores muestrales obtenidos

con los de la población desconocida; juzga así, la probabilidad del parámetro desconocido (al cual se refiere la hipótesis) en función de los datos obtenidos, que esperan sean similares al valor de dicho parámetro.

Por otra parte, no es posible mantener, con los resultados obtenidos, como causa principal de los errores de interpretación que afectan al nivel de significación los problemas derivados de una insuficiente comprensión del concepto de probabilidad condicional. Nuestros resultados avalan la hipótesis de que es la concepción del contraste como *prueba probabilística de la hipótesis*, (Vallecillos, 1995), predominante entre los alumnos entrevistados, la que está condicionando la interpretación que hacen del nivel de significación. En efecto, este es considerado como la probabilidad a posteriori de la hipótesis nula y, por tanto, se interpreta como  $P(H_0 \text{ sea cierta} \mid H_0 \text{ ha sido rechazada})$ , expresión en donde aparecen intercambiados los dos sucesos que intervienen en la definición del nivel de significación.

Finalmente queremos también resaltar el modo en que diversos errores conceptuales han influido en los procedimientos de los alumnos, en diversas partes de la prueba:

- La confusión entre hipótesis nula (alternativa) y región de aceptación (rechazo) ha llevado a los alumnos a plantear unas hipótesis incorrectas en el apartado a) del Problema 1.
- La creencia en la posibilidad del cálculo de la probabilidad de la hipótesis ha llevado a tomar el valor de la probabilidad asociada a la región crítica como probabilidad de la hipótesis nula.
- La falta de consideración de la distribución muestral del estadístico ha hecho emplear en todos los cálculos la distribución original de la población, produciendo errores en la tipificación de la variable y valores anormalmente grandes para el nivel de significación que, sin embargo, no han llamado la atención de los alumnos.
- La concepción errónea sobre el nivel de significación ha bloqueado el cálculo del nivel de significación, debido a la creencia en que éste valor no se calcula, sino que se fija al comienzo del proceso.

Vemos en todos estos casos la gran dificultad que supone en la resolución de estos problemas tener que emplear tan gran número de conceptos y procedimientos estadísticos cuyo conocimiento se supone adquirido. Como señalan Hiebert y Lefevre (1987), cuando los conceptos y procedimientos no están conectados, los estudiantes pueden dar respuestas a los problemas sin comprender lo que hacen. Vemos también la separación que se produce en la enseñanza entre los diferentes conceptos que se enseñan aislados al alumno, muchas veces sin mostrar su necesidad. En el cálculo del nivel de significación de un contraste, en el contexto de control de calidad u otro semejante al propuesto, hay una ocasión excelente de mostrar al alumno la utilidad del cálculo de probabilidades, con la ayuda de la distribución normal. Sin embargo, al llegar a este tema, se supone que ello es una habilidad adquirida, con lo que el alumno no llega a realizar este tipo de ejercicios. En consecuencia, esperamos que estos resultados proporcionen información a los profesores sobre las dificultades de sus alumnos y sobre los puntos concretos en los cuales la enseñanza debe prestar atención cuidadosa. Al fin y al cabo sólo la mejor enseñanza de los test de hipótesis traerá consigo la mejora en sus aplicaciones futuras.

*Agradecimientos:* Este trabajo se realiza con financiación del Proyecto de Investigación PS93-0196 de la DGICYT, M.E.C. (Madrid). Agradecemos también a los estudiantes que han participado su meritoria labor y su esfuerzo.

## 5. REFERENCIAS

- Artigue, M. (1990). Epistemologie et didactique. *Recherches en Didactique des Mathématiques*, 10(2/3), 241-286.
- Bakan, D. (1966). The tests of significance in psychological research. *Psychological Bulletin*, 66, 423-437.
- Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics*, 4, 24-27.
- Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis Doctoral. Universidad

de Granada.

- Diaconis, P. y Freedman, D. (1981), The persistence of cognitive illusions. *The Behavioural and Brain Sciences*, 4, 333-334.
- 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. *Theory and Psychology*, 5 (1), 75-98.
- Godino, J. D. y Batanero, C. (1994), Significado personal e institucional de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, 14(3), 325-355.
- Hawkins, A., Jolliffe, F. y Glickman, L. (1992), *Teaching Statistical Concepts*. New York: Longman.
- Hiebert, J. y Lefevre, P. (1987). Conceptual and procedural knowledge in mathematics: an introductory analysis. En J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics*. London: LEA.
- Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.
- Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal*, 5(1), 4-18.
- Morrison, D. E. y Henkel, R. E. (Ed.). (1970), *The significance tests controversy - a reader*. Chicago: Aldine.
- Schuyten, G. (1991), Statistical thinking in psychology and education. *Proceedings of the II International Conference on Teaching Statistics*, (vol.2, pp.496-489). Dunedin: Universidad de Otago.
- Scott, P. (1988), *Introducción a la investigación y evaluación educativa. Lecturas en Educación Matemática*. México: UNAM.
- Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: three suggested reforms. *Educational Researcher*, 25(2), 26-30.
- 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. (1995). Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios. *Recherches en Didactique des Mathématiques*, 15, (3), 53-81.
- Vallecillos, A., Batanero, C. y Godino, J. (1992). Student's understanding of the significance level on statistical tests. En W. Geesling y K. Graham (Eds.), *Proceedings of the XVII Conference on the Psychology of Mathematics Education*, (vol.1, pp. 49-56). University of New Hampshire.
- Vallecillos, A. y Batanero, C. (1996), Conditional probability and the level of significance in tests of hypotheses. En L. Puig y A. Gutiérrez (Eds.), *Proceedings of the XX PME Conference on the Psychology of Mathematics Education*. (vol.4, pp. 271-378). Universidad de Valencia.
- Vergnaud, G. (1982), Cognitive and developmental psychology and research in mathematics education: some theoretical and methodological issues. *For the Learning of Mathematics*, 3(2), 31-41.
- Vergnaud, G. (1990), La theorie des champs conceptuels. *Recherches en Didactique des Mathématiques*, 10,(2-3), 133-170.
- Zacks, S. (1981), *Parametric statistical inference. basic theory and modern approaches*. Oxford:Pergamon Press.