

El Papel de las Hipótesis Estadísticas en los Contrastes: Concepciones y Dificultades de Aprendizaje

Resumen

En este artículo se recogen los resultados que se refieren al concepto de hipótesis estadística obtenidos en una investigación experimental más general llevada a cabo con estudiantes universitarios (Vallecillos, 1994). En ellos se ponen de manifiesto una serie de dificultades y errores que dificultan o impiden el aprendizaje del tema del contraste de hipótesis y sus aplicaciones. En particular, se ha prestado atención a dos aspectos: a) la influencia que los errores que afectan al establecimiento de las hipótesis han tenido sobre la resolución del problema propuesto y b) la forma en que las concepciones sobre las hipótesis han influenciado las concepciones de los estudiantes sobre la lógica global del proceso de contraste. Se han caracterizado las principales concepciones sobre el concepto de hipótesis que consideramos información de interés para los profesores implicados en la enseñanza del tema.

Abstract: In this article are collected the results concerning statistical hypothesis concept obtained in an experimental investigation more general carried out with undergraduates (Vallecillos, 1994). In them a series of difficulties and mistakes that render difficult or hinder the learning of the hypothesis testing process and its applications are presented. In particular, it has been paid attention to two aspects: a) the influence that the mistakes affecting to the establishment of the hypothesis have had on the resolution of the proposed problem and b) the form in which the conceptions on the hypothesis have influenced the students' conceptions on the global logic of the hypothesis testing process. The principal conceptions on the hypothesis concept that we consider interesting information for the teachers implicate in the teaching of the topic have been characterized.

Angustias Vallecillos Jiménez

Universidad de Granada, España

Pedagogía (43), Psicología (70), Medicina (61), Matemáticas (31) y Empresariales (75). Son estudiantes de 2º, 4º y 5º cursos 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.

Cuestionario empleado:

ítem 1. ¿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:

Se refiere a la idea de hipótesis paramétrica, sobre un valor del parámetro de la distribución de la variable en la población. Se trata de estudiar además si el estudiante distingue correctamente el estadístico de contraste del parámetro poblacional y si conoce la notación habitual.

En los cursos previos a la universidad, los estudiantes han trabajado la estadística descriptiva fundamentalmente y no se enfatiza quizás suficientemente la diferencia entre los dos conceptos de parámetro poblacional y estadístico muestral que sólo tiene interés desde el punto de vista de la realización de inferencias de la muestra a la población. Los estudiantes pueden no captar la idea del estadístico como realización de una variable aleatoria.

Los dos primeros distractores A y B plantean la posibilidad de que la hipótesis nula pueda referirse a un único valor del estadístico, que algunos estudiantes identifican sólo con la media muestral. Algunos otros identifican la expresión «hipótesis nula» con el hecho de que el único valor admisible para la hipótesis sea $\mu_x = 0$ o $\sigma_x = 0$.

En D se plantea la posibilidad de que la hipótesis deba ser siempre un único valor numérico, esto es, una hipótesis nula simple. En este distractor se presenta una hipótesis relativa a la igualdad de medias en dos poblaciones para estudiar también si podría creerse que la hipótesis nula se refiere a una única población.

Se pide razonar la respuesta con el fin de profundizar más en ella y comprobar la estabilidad de la misma.

ítem 2. En una encuesta electoral se desea averiguar si hay el mismo número de americanos a favor de la política económica del Presidente que en contra de la misma. Supuesto que p representa la probabilidad asociada a los habitantes que están de acuerdo con dicha política económica y $q = 1 - p$, ¿Cuál de las siguientes hipótesis elegirías como hipótesis nula?

- A: $p > q$.
- * B: $p = q = 1/2$.
- C: $p \neq q \neq 1/2$.
- D: $q > p$.

Analizamos la idea de hipótesis estadística que tienen los estudiantes, desde un punto de vista teórico, en un contexto de aplicación. A partir del enunciado el estudiante debe elegir la hipótesis nula más adecuada, de entre unas opciones dadas. Está subyacente la distinción entre hipótesis nula y alternativa que es una dificultad para los estudiantes ya detectada con anterioridad (Peskun, 1987; Batanero y cols., 1994).

Este ítem está muy relacionado con la lógica global del proceso ya que la elección de las hipótesis nula y alternativa está condicionada por la concepción del contraste que se tenga y el papel que en él se asigne a las hipótesis. Puesto que en las aplicaciones de los contrastes se trata normalmente de 'confirmar' que un determinado tratamiento novedoso (médico, educativo, etc.) tiene efecto real sobre los sujetos, la hipótesis nula es la hipótesis de que no hay diferencias entre los tratamientos, esto es, la que se desea rechazar. La hipótesis alternativa es la que se desea confirmar. Los distractores empleados plantean distintas posibilidades de confusión entre las hipótesis nula y alternativa o tomar como hipótesis nula una hipótesis de investigación que se desea confirmar.

Análisis de los Resultados: El papel de las Hipótesis en los Contrastes

En primer lugar se han analizado los resultados del cuestionario en ambos ítems. A continuación se ha analizado conjuntamente la respuesta al ítem 1 con las argumentaciones escritas por los estudiantes en su apoyo. Por último se analizan los resultados obtenidos en cuanto al establecimiento de las hipótesis adecuadas a la situación en el problema propuesto.

Análisis de la respuesta a los ítems

Los resultados obtenidos en los ítems analizados están recogidos en la Tabla 1 a continuación.

Un 56% de las respuestas obtenidas en el ítem 1 son correctas. De ellas el mayor porcentaje, 90.3%, corresponde a los estudiantes de Matemáticas, si bien los grupos de estudiantes de Psicología y Medicina son también muy altos, 77.1% y 73.8%, respectivamente.

El distractor D ha sido elegido por un porcentaje moderadamente alto de estudiantes, 18.5%. Esta opción refleja un error que lleva a considerar a muchos estudiantes que una hipótesis debe venir expresada por un valor numérico. La forma en que está expresada a hipótesis ha resultado conceptualmente difícil para los estudiantes que, en general, no han sabido ver una forma equivalente más sencilla para esta hipótesis: $\mu_1 - \mu_2 = 0$. El porcentaje más alto en esta opción, por especialidades, corresponde a los estudiantes de Informática con un 34.9%.

Los otros distractores, asociados al error de considerar las hipótesis referidas a un único valor del estadístico muestral han sido elegidos por un bajo porcentaje de estudiantes.

En este ítem la notación juega un papel esencial: algunos de los errores detectados pueden ser debidos a errores de notación y no conceptuales. Para dar respuesta a este interrogante hemos analizado los argumentos dados por los estudiantes para explicar su respuesta al ítem.

Tabla 1: Frecuencias y porcentajes de respuestas

Opción	ítem 1 (%)	ítem 2 (%)
A	3(0.7)	58(13.3)
B	34(7.8)	*299(68.6)
C	*244(56.0)	28(6.4)
D	81(18.5)	24(5.5)
Blanco	74(17.0)	27(6.2)
Total	436(100.0)	436(100.0)

Nota: con * la respuesta correcta

El porcentaje de respuestas en blanco encontradas en este ítem, 17%, creemos que refleja bastante bien la dificultad de los estudiantes para reconocer una posible hipótesis y distinguirla de otra que no lo es.

Un 68.6% de las respuestas obtenidas en el ítem 2 son correctas. Aquí el porcentaje mayor de respuestas correctas ha correspondido al grupo de estudiantes de Medicina con un 91.8%, seguido de los grupos de Matemáticas y Empresariales con un 90.3% y 84%, respectivamente.

El distractor A tiene un porcentaje de respuesta mayor, en general, que los demás distractores, si bien es bajo en todas las especialidades. Está asociado al error que consiste en tomar como hipótesis nula aquella que se desea confirmar en el sentido siguiente: mientras que una hipótesis de investigación está, normalmente, redactada en sentido positivo, en términos de expectativas en los resultados (Scott, 1988), las hipótesis nulas se plantean en términos de no diferencias, de ineffectividad del tratamiento. En el caso planteado algunos estudiantes pueden haber asociado las expectativas de ganancia electoral con el establecimiento de la hipótesis nula en un particular proceso de contraste porque, como afirman Hawkins y cols. (1992), las intuiciones de los estudiantes les llevan

a buscar resultados confirmatorios más que falsatorios de las hipótesis. Este error nos está indicando por lo tanto, además, una confusión entre ambos tipos de hipótesis, de investigación y estadísticas en estos alumnos. Es el único error encontrado en la especialidad de Matemáticas, 9.7%, en donde no hay tampoco ninguna respuesta en blanco.

En general, los porcentajes de respuesta obtenidos en las demás opciones son bajos en todas las especialidades, lo que indica que no hay diferencias apreciables al respecto entre los distintos grupos de estudiantes. Tampoco en lo que se refiere a las respuestas en blanco.

Análisis de argumentos en el ítem 1

Los argumentos que han dado los estudiantes para apoyar sus respuestas a la opción elegida del ítem han sido sometidas a un riguroso análisis de contenido (Bardin, 1986) y clasificadas en categorías de análisis según que sean correctas o no y, en este caso, agrupando en una misma categoría respuestas conceptualmente homogéneas según los errores señalados por los distractores empleados, principalmente.

La clasificación obtenida se recoge a continuación. En cada categoría citamos una respuesta textual de un estudiante para ilustrar mejor la clasificación efectuada.

1. *Argumento correcto*: porque la hipótesis nula debe referirse a la población y no a la muestra.

- "No es una H_0 porque las hipótesis se establecen sobre los parámetros de la población y no sobre los estadísticos de la muestra".

2. *Argumento correcto*: porque la hipótesis nula debe referirse a la población y no a la muestra.

- "Se debe a que 35 es el valor de una muestra y se consideran los poblacionales".

Estos dos argumentos correctos se han mantenido como categorías de análisis distintas por varias razones. En primer lugar presentan frecuencias grandes de aparición. Por otra parte, nos ha parecido relevante, a efectos didácticos, la referencia explícita al parámetro poblacional o al estadístico muestral que hacen algunos estudiantes. Desde el punto de vista de la Didáctica de la Estadística, es preciso tener en cuenta el distinto grado de dificultad que pueden tener para los estudiantes estos dos conceptos: en el primer caso se trata de una característica de una variable aleatoria en la población de referencia mientras que en el segundo se trata de un estimador de la misma en la muestra. La notación empleada en ambos casos es distinta y esto exige también su diferenciación. Schuyten (1991) señala la gran dificultad que supone para los estudiantes el hecho de que un mismo concepto —por ejemplo el concepto de media— se emplee en inferencia estadística a distintos niveles de abstracción.

3. *Argumento parcialmente correcto*: porque es un dato y no una hipótesis.

Las respuestas clasificadas aquí son esencialmente correctas en lo que se refiere al contenido del ítem, si bien algunas de ellas hacen referencias explícitas incorrectas a otras cuestiones que aconsejan su inclusión como parcialmente correctas, o bien son incompletas.

- "Todas las demás son hipótesis sobre la población sobre la cual tenemos cierto conocimiento; pero C se refiere al valor de la media muestral, pero esto no es una hipótesis es algo que conocemos una vez conocida la muestra o elegida la muestra, es algo que conocemos siempre".

4. *Argumento erróneo*: porque la hipótesis nula tiene que ser un valor numérico.

- "Porque la hipótesis nula no nos da un valor numérico".

(Respuesta al ítem: $D: \mu_1 = \mu_2$).

5. *Argumentos erróneos* basados en que la hipótesis debe ser un estadístico muestral y no puede ser un número.

- "Nunca se conoce el valor de σ_x^2 . Se puede decir que $H_0: \sigma_x^2 \neq \sigma_y$ o $\sigma_x > 0 \dots$ pero nunca que tiene un valor concreto".

- "Porque no es un estadístico".

6. *Argumentos confusos o difícilmente interpretables*.

En la Tabla 2 están recogidas las frecuencias y porcentajes de respuesta obtenidos. Los porcentajes incluidos en ella se han calculado sobre las respuestas efectivamente obtenidas, excluyendo las repuestas en blanco, ya que al ser muy numerosas, 38.8%, enmascararían los resultados reales obtenidos.

El hecho de que sólo hayan argumentado su respuesta un 61.2% de los estudiantes en un dato que refleja la dificultad generalizada que tienen, en todas las especialidades, para establecer con claridad una hipótesis nula y distinguirla de una que no lo es. En menor medida, esta dificultad también se ha reflejado en la respuesta al ítem.

Si analizamos las argumentaciones obtenidas, el mayor porcentaje de respuestas correctas corresponde al segundo argumento correcto que no hace referencia explícita a la diferenciación entre el estadístico muestral y el parámetro poblacional. Este dato lo consideramos de interés porque se suma a otros indicadores de que los estudiantes no son conscientes del papel jugado por el estadístico, como variable aleatoria, en el proceso de contraste. En total hay un 58.4% de respuestas correctas. Si sumamos también las parcialmente correctas llegamos a un 73.8% de argumentaciones que podemos considerar correctas. Esto es, si tenemos en cuenta sólo a los estudiantes que han argumentado su respuesta, el porcentaje de ellos que son capaces de distinguir un hipótesis nula planteada de otra que no lo es, es muy alto, en general en todos los grupos de estudiantes.

Los porcentajes de argumentos erróneos no llegan en ningún caso al 10% y son bastante homogéneos en las tres categorías por lo que podemos considerarlos de poco interés.

Tabla 2: Frecuencias y porcentajes de argumentos en el ítem 1

Argumentos		Frecuencia (%)
Porque la hipótesis nula debe referirse a un parámetro poblacional y no a un estadístico muestral	1 Correcto	74 (27.7)
Porque la hipótesis debe referirse a la población y no a la muestra	2 Correcto	82 (30.7)
Porque es un dato y no una hipótesis	3 Parc. Corr.	41 (15.4)
Porque la hipótesis nula debe ser un dato numérico	4 Erróneo	25 (9.4)
Porque la hipótesis nula tiene que ser un estadístico muestral	5 Erróneo	23 (8.6)
Argumentos confusos o inoportunos	6 Erróneo	22 (8.2)
Total		267(100.0)

Análisis conjunto de respuestas y argumentos en el ítem 1

Una vez analizadas las respuestas y las argumentaciones dadas por los estudiantes hemos hecho un análisis conjunto de ambos tipos de respuestas para estudiar la estabilidad y consistencia de la respuesta a la opción del ítem elegida. Para ello hemos agrupado en una sola opción todas las respuestas incorrectas del ítem de manera que sólo quedan dos opciones, correcta y no correcta. Los resultados obtenidos se recogen en la Tabla 3.

El total de los estudiantes que han respondido el ítem y se han apoyado en una argumentación correcta asciende al 58.4%. El argumento mayoritariamente empleado por ellos es el codificado como 2, utilizado por el 30.7% de los estudiantes.

Un 98.6% de los estudiantes que han dado el primer argumento correcto también han elegido la opción correcta del ítem. Por el contrario, sólo un 1.4% de los que han respondido incorrectamente el ítem han utilizado este argumento correcto.

Los porcentajes correspondientes al segundo argumento correcto son también similares: 95.1% y 5.4%, respectivamente.