

PRUEBA DE HIPÓTESIS ESTADÍSTICA. ESTUDIO DE DIFICULTADES CONCEPTUALES EN ESTUDIANTES DE GRADO Y DE POSTGRADO

María Inés Rodríguez, José Armando Albert Huerta

UNIVERSIDAD NACIONAL DE RÍO CUARTO (ARGENTINA) -INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE
MONTERREY (MÉXICO)

mrodriguez@exa.unrc.edu.ar, albert@itesm.mx

Resumen. La práctica de consultoría estadística nos ha revelado algunas dificultades y errores de investigadores de distintas disciplinas en la utilización de las pruebas de hipótesis estadísticas, particularmente, en la interpretación de resultados. Estas dificultades pudieran tener origen y solución desde la formación universitaria, por lo que nuestro propósito es describir la naturaleza de esas dificultades en el contexto de estudiantes universitarios e investigadores sudamericanos. Aunque el problema que plantean esas dificultades y su prevención puede estudiarse desde la complejidad de sus componentes sistémicas, esta investigación se centra principalmente en la epistemológica y en la cognitiva. Se realizó un estudio exploratorio aplicando una encuesta, adaptada de la utilizada por Vallecillos (1996), a una muestra de estudiantes de grado y a otra de postgrado. Se analizan en este trabajo los resultados relativos al nivel de significación.

Palabras Clave: Inferencia estadística. Pruebas de significación. Nivel de significación.

Introducción

Uno de los problemas principales en un curso introductorio de estadística a nivel universitario es hacer la transición del análisis de datos a la inferencia (Moore, 1998a). Los dos tipos más corrientes de inferencia estadística son los intervalos de confianza, utilizados para estimar el valor de un parámetro poblacional y los test o pruebas de hipótesis, que valoran la evidencia a favor de una determinada afirmación sobre una población. Los dos tipos de inferencia establecen las probabilidades de lo que ocurriría si se utiliza el método de inferencia muchas veces. Esta investigación trata del estudio de las pruebas de hipótesis

(o de significación según Fisher) por dos razones: una de tipo didáctico, señalada por varios investigadores (Vallecillos, 1996; Batanero, 2000) y que compartimos, quienes han detectado algunos obstáculos en la adquisición de conceptos relativos a los tests en alumnos universitarios; la otra es que en la actualidad la metodología inferencial es muy utilizada por la mayoría de los investigadores de todas las disciplinas. Sin embargo, se ha encontrado que suelen haber grandes confusiones conceptuales y procedimentales en su aplicación, que traen aparejados grandes riesgos en sus conclusiones. Al respecto, Falk y Greenbaum (1995) señalan: “Para ser justos, el hecho de que la gente malinterprete los tests de significación no es una falta de los tests, y no hay razón para descartarlos en conjunto. Las malas interpretaciones pueden, después de todo, ser clarificadas y restablecido el significado correcto.[...] Sugerimos que, a menos que sean tomadas fuertes medidas en la enseñanza estadística, las probabilidades de que se remuevan estas concepciones erróneas son bajas en el presente.[...] Se requiere un esfuerzo educacional masivo para erradicar esta concepción errónea”, (p. 98).

La importancia y preocupación vigente del tema, dentro de la comunidad de investigadores en educación estadística, lo demuestra la realización en Agosto del corriente año, del 5^º *Foro sobre el Razonamiento, Pensamiento y Alfabetización Estadística*, (SRTL-5), cuyo tema central de debate y análisis fue, “Razonamiento acerca de la Inferencia Estadística: Maneras innovativas de conectar la probabilidad y los datos”, que tuvo lugar en la Universidad de Warwick, Coventry. (R.U.). Este Foro se realiza desde 1999, cada dos años sobre diversos tópicos de interés a partir de las problemáticas planteadas tanto por investigadores como por docentes de estadística, en los dos últimos foros (SRTL-3 y 4). Algunos de los temas sugeridos para la presentación de trabajos en este SRTL-5 fueron:

- ¿Cuáles son las ideas más simples de Inferencia Estadística que pueden comprender los estudiantes?

- ¿Cómo se desarrolla el Razonamiento Estadístico Inferencial desde su manera más simple (informal) a una más compleja (formal)?
- ¿Qué tipo de errores se encuentran en los estudiantes acerca de la Inferencia Estadística?

Esta serie de interrogantes, planteadas por el SRTL-5, pone en evidencia que la preocupación sobre la problemática de la enseñanza de la Inferencia Estadística en general existe en distintas latitudes y ha motivado la realización de la presente investigación.

Respecto a la evidencia del uso, muchas veces incorrecto de esta temática, Falk y Greenbaum (citado en Batanero, 2000) sostienen: “a pesar de las recomendaciones, los investigadores experimentales persisten en apoyarse demasiado en la significación estadística, sin tener en cuenta los argumentos de que los tests estadísticos por sí solos no justifican suficientemente el conocimiento científico. Algunas explicaciones de esta persistencia incluyen la inercia, confusión conceptual, falta de mejores instrumentos alternativos o mecanismos psicológicos, como la generalización inadecuada del razonamiento en lógica deductiva al razonamiento en la inferencia bajo incertidumbre” (p. 2).

El problema

La práctica de consultoría estadística nos ha revelado algunas dificultades y errores de investigadores de distintas disciplinas en la utilización de las pruebas de hipótesis estadísticas, particularmente en la interpretación de resultados. Esta situación es coincidente con la señalada por Lécoute (2006). El proceso de inferencia sólo es confiable cuando los datos de partida se han obtenido de un experimento correctamente aleatorizado. Es decir, cuando se emplean los métodos inferenciales se considera que los datos provienen de una muestra aleatoria o de un diseño experimental correspondientemente aleatorizado. Si esto no ocurre, las conclusiones pueden ser un

desafío al sentido común o a la verdad. Esta idea básica y fundamental no es fácilmente comprendida por los estudiantes, como tampoco por gran cantidad de investigadores que utilizan la inferencia estadística para el análisis científico de sus datos experimentales como una metodología de evaluación objetiva de su hipótesis en cuestión.

Por otra parte, la lógica de las pruebas de hipótesis presenta dificultades conceptuales vinculadas a la filosofía y a la psicología que las hacen susceptibles de interpretaciones incorrectas. Del desarrollo conceptual de Pruebas de Hipótesis surge una controversia, que llega hasta nuestros días aún no resuelta. Ito (1999) describe tres niveles diferentes que interactúan en la controversia respecto a los contrastes estadísticos:

- (a) La disputa dentro de la misma estadística, donde diferentes métodos e interpretaciones varias de los mismos métodos fueron recomendados por los enfoques de Fisher, Neyman-Pearson y el Bayesiano.*
- (b) La controversia en la aplicación de la estadística donde, en la práctica, el contraste de significación es una mezcla informal de los contrastes de significación originales de Fisher, la teoría de Neyman-Pearson y conceptos e interpretaciones que no son partes de esta última. Mas aún, los editores de revistas y las sociedades profesionales están sugiriendo cambios en las políticas de publicación científica, respecto a los métodos estadísticos.*
- (c) La controversia en la enseñanza acerca de cuándo, cómo y con qué profundidad deberíamos enseñar la inferencia estadística.*

Estos tres niveles están de hecho interrelacionados y afectan a investigadores, profesores y alumnos en sus concepciones sobre Estadística. El propósito de esta investigación es describir la naturaleza de las dificultades en el contexto de estudiantes universitarios e investigadores sudamericanos. Para ello, se realizó un estudio exploratorio aplicando una encuesta a una muestra de estudiantes de grado y a otra de postgrado.

Aunque el problema puede estudiarse desde la complejidad de sus componentes sistémicas —la epistemológica, cognitiva, didáctica y de prácticas sociales—, esta investigación se centra principalmente en la epistemológica y en la cognitiva.

Diferencias y similitudes entre los enfoques de las pruebas de hipótesis

Para contribuir a una mejor interpretación y justificación de nuestro trabajo, incluimos a continuación una breve síntesis del enfoque de cada una de las escuelas que dieron origen a la teoría de los tests de hipótesis. Consideramos importante aclarar y difundir los diferentes enfoques, debido a que en la actualidad una mezcla de la lógica sugerida por estos distintos enfoques es utilizada por la mayor parte de los investigadores en sus prácticas y por los educadores en su actividad de enseñanza. Generalmente esto ocurre por desconocimiento de sus diferencias, ya que las mismas no radican en los cálculos algorítmicos, sino más bien en el razonamiento subyacente. Por otra parte, la mayoría de los libros de texto optan por un solo enfoque y son pocos los que hacen explícitas las ideas sostenidas por los diferentes autores. Entre éstos, podemos citar el de D. Moore (1998a), que si bien desarrolla el tema desde la teoría de Fischer aduciendo que es la utilizada por la mayoría de los usuarios de la estadística, también alude a las ideas de Neyman y Pearson y cita la referencia para conocer el tema desde ese enfoque. Otro ejemplo es el texto de Gómez Villegas (2005), que presenta todos los temas desde dos perspectivas, la de los métodos clásicos y la bayesiana, sin ningún menoscabo de uno frente a otro, construyendo los contrastes de hipótesis bajo el enfoque de Neyman y Pearson.

Cabe señalar también que las notas de clase de la mayoría de los cursos básicos de estadística, utilizados en la Universidad Nacional de Río Cuarto (UNRC), como así también en universidades de otras latitudes, que hemos encontrado en Internet, no hacen referencia alguna al desarrollo histórico del tema, ni a las motivaciones que provocaron su surgimiento, que en gran medida están vinculadas al desarrollo de otras disciplinas originando así concepciones y terminologías que fueron causa de discusiones entre sus autores no superadas hasta el momento.

▪ **Las pruebas de significación de Fisher**

La clave de los tests Fisherianos es que no hacen referencia a la hipótesis alternativa. Realmente pueden ser considerados como un procedimiento de validación de un modelo. Se tiene la distribución del modelo (planteada en la hipótesis nula) y se examina si los datos parecen raros o adecuados a ese modelo planteado. El nivel de significación α es simplemente una regla de decisión por la cual se rechaza o no la hipótesis nula. En otras palabras, es un valor de probabilidad que indica cuán raros deben ser los datos para rechazar la hipótesis nula. Implícitamente el nivel α determina qué datos llevarían al rechazo de la hipótesis nula y cuáles no (Christensen, 2005). En este enfoque el valor p es realmente un concepto más importante que el valor α . Cuando $p < \alpha$ se dice que los resultados son estadísticamente significativos.

Para redondear este enfoque, podemos decir que para Fisher los tests de significación, sirven para confrontar la hipótesis nula con los datos observados, siendo el valor p de la prueba el que indica la fuerza de la evidencia en contra de la hipótesis nula. Según su punto de vista, las pruebas estadísticas no deben ser utilizadas para realizar inferencias inductivas de la muestra a la población, sino más bien, una inferencia deductiva de la población de las posibles muestras a la muestra particular obtenida en cada caso. Para él habría que tener en cuenta que la hipótesis nula nunca resulta probada en el curso de un experimento, pero sí posiblemente refutada. Podemos señalar como características de las pruebas de significación estadística, según Fisher, los siguientes puntos:

- El razonamiento inductivo no puede confirmar una hipótesis.
- Un contraste sólo sirve para rechazar una hipótesis, nunca para confirmarla.
- No existen errores de tipo II, esto es, no se considera el error que se comete al aceptar una hipótesis nula, bajo el supuesto que sea falsa.

- **El contraste de hipótesis como proceso de decisión (Teoría de Neyman- Pearson)**

Mientras que en el enfoque anterior la prueba se utiliza para evaluar la fuerza de la evidencia en contra de una hipótesis, para Neyman y Pearson (N-P) el problema de una prueba de hipótesis estadística aparece en circunstancias en que estamos forzados a realizar la elección entre dos cursos de acción. Esta dupla defendió una versión denominada contraste de hipótesis que mezcla la idea de prueba de significación y la idea de regla para tomar decisiones. Con N-P nace la idea de que es necesario considerar otra hipótesis, surgiendo así la hipótesis alternativa, con el origen del concepto de test uniformemente más potente. Para ellos una prueba de hipótesis estadística es una regla de comportamiento inductivo.

En este enfoque se plantea lo siguiente: considerando una muestra aleatoria (x_1, \dots, x_n) de una población X con función de densidad f_θ , donde $\theta \in \Theta$, se trata de decidir con base en la información que brinda esta muestra si es posible que la misma provenga del espacio paramétrico Θ_0 o a otro espacio Θ_1 , siendo Θ_0 y Θ_1 una partición previamente elegida del espacio paramétrico Θ . Dicho en lenguaje estadístico, queremos contrastar la hipótesis nula $H_0 : \theta \in \Theta_0$ frente a la hipótesis alternativa $H_1: \theta \in \Theta_1$, para decidir con la evidencia que brinda la muestra, cuál de ellas es válida. Por esta razón se debe colocar como hipótesis nula aquélla sobre la que se tenga mayor convencimiento de que va a ser cierta, ya que estas pruebas se construyen tratando de proteger a la hipótesis nula. Muy por el contrario, en los contrastes bayesianos esta consideración no se tiene en cuenta. Resumiendo, los rasgos característicos del enfoque de Neyman-Pearson son los siguientes:

- La inducción puede proporcionar un criterio para aceptar una hipótesis.
- Se reconocen dos tipos de errores: el de tipo I y el de tipo II.
- Se trata de buscar el test con alto valor de potencia $(1 - \beta)$.
- Se da una interpretación a la hipótesis alternativa.

Es importante resaltar que la probabilidad de que la hipótesis nula sea cierta una vez que la hemos rechazado y la probabilidad de que la hipótesis nula sea cierta una vez que hemos obtenido los datos, no pueden conocerse a partir del contraste de hipótesis, en las acepciones de Fisher o de Neyman-Pearson. El único enfoque que permite calcular estas probabilidades a posteriori de la hipótesis es el bayesiano, por el cual el experimentador puede revisar su grado de aprobación de la hipótesis, en vista de los resultados obtenidos.

- **Enfoque bayesiano de los test**

Los fundamentos de la aproximación bayesiana a la inferencia están basados en los trabajos de Thomas Bayes (1701- 1761), cuya publicación póstuma fue realizada por Price en 1763, bajo el título *Un ensayo hacia la resolución de un problema en la doctrina del azar*. La aproximación bayesiana al contraste de hipótesis continuó desarrollándose hasta la aparición de los trabajos de E. Pearson, J. Neyman y R. Fisher, que interrumpieron la utilización de los métodos bayesianos, básicamente por depender de la distribución inicial, lo que hace que distintas personas con diferente información inicial y la misma información muestral lleguen a conclusiones distintas. Es decir, ninguno de los autores citados utilizaron el teorema de Bayes de la probabilidad condicional como base para la inferencia estadística, y particularmente Fisher era reacio a ello. Continuaron con la aproximación bayesiana, entre otros, Ramsey con su publicación *Fundamentos de Matemáticas*, en 1931; Jeffreys, con su *Teoría de la Probabilidad* en 1939; De Finetti, 1974; Lindley en 1975, quienes han promovido que actualmente los métodos bayesianos, en particular los contrastes de hipótesis, presenten un elevado desarrollo, incorporando la información a priori para resolver el problema de la asignación de probabilidades de la hipótesis.

El enfoque bayesiano considera el parámetro μ como una cantidad aleatoria con una distribución de probabilidad conocida que expresa la información imprecisa que se tiene acerca de su valor. Por ejemplo, la altura media μ de todas las estudiantes de la UNRC no es aleatoria en el sentido tradicional, pero es incierta. En la perspectiva bayesiana, el concepto

de probabilidad se amplía para incluir estas *probabilidades subjetivas* o personales. Lo que es nuevo aquí no son las matemáticas, que permanecen iguales, sino la interpretación de la probabilidad como representación de una estimación subjetiva, de la incertidumbre en vez de considerarla una frecuencia relativa en el largo plazo, como sostiene Moore (1998b) para quien la conclusión bayesiana es sin lugar a dudas más sencilla de entender que el enunciado clásico (cap. IV, p. 24).

Las características de esta escuela bayesiana se resumen en dos:

- Por un lado utilizan el teorema de Bayes como base para la inferencia estadística.
- Por otro lado, y donde la controversia es mayor, asigna probabilidades a una hipótesis.

Se puede observar que hay diferencias conceptuales respecto a los contrastes de hipótesis entre las distintas escuelas, pero ellas, por lo general, no se tienen en cuenta en el proceso de su enseñanza. Frecuentemente este tema se imparte con una mezcla de los tres enfoques. Por ejemplo, se señalan las dos hipótesis y los dos tipos de error tanto cuando se explica este método para aplicarlo como regla de decisión, como cuando se enseña en cursos orientados a las ciencias experimentales donde comúnmente sólo interesa encontrar diferencias estadísticamente significativas, sin importar la magnitud del efecto. “Libros de texto como el de Guilford (1942) contribuyeron a difundir una mezcla de las lógicas de los contrastes de significación de Fisher con algunos componentes de Neyman-Pearson, dando una interpretación Bayesiana al nivel de significación y otros conceptos relacionados”. (Batanero, 2000, p.13)

Metodología

Con el propósito de tener un primer acercamiento a las dificultades de comprensión del tema de prueba de hipótesis, de estudiantes e investigadores universitarios, se realizó un estudio exploratorio aplicando una encuesta a una muestra de estudiantes de grado y a otra

de postgrado. La encuesta fue una adaptación de la utilizada por Vallecillos (1996), cuya validez y fiabilidad fueron por ella comprobadas, la cual consta de dos partes. Un cuestionario sobre contenidos conceptuales, al que se le incorporaron dos preguntas referidas al valor p de una prueba, que Vallecillos no había considerado y otra parte referida a conocimientos procedimentales con base en una situación problema. A continuación se muestran las preguntas de mayor dificultad vinculadas al nivel de significación y a sus elementos relacionados, tales como interpretación, determinación y definición del mismo.

- **Interpretación del nivel de significación:**

Su análisis se efectuó mediante los ítems 3 y 6 de la encuesta, que a continuación se presentan:

Item 3: Supongamos que el nivel de significación es $\alpha = 0.01$. Esto significa que:

- A: Si la hipótesis es cierta, tendremos una probabilidad 0.01 de rechazarla.
- B: Si la hipótesis nula es falsa, tendremos una probabilidad 0.01 de aceptarla.
- C: Hay una probabilidad 0.01 de que nuestro resultado sea correcto.
- D: Si hemos decidido rechazar la hipótesis nula, tenemos una probabilidad de error del 1%.

Se estudia con este ítem la probabilidad de error de tipo I y su posible confusión con el error de tipo II. También se estudia la confusión de los condicionales en la definición del nivel de significación. (Se resalta siempre con negrita la respuesta correcta).

Los distractores presentados para analizar la consistencia de las respuestas obtenidas son:

- B: Interpretación del nivel de significación en términos de error de tipo I y su posible confusión con el error de tipo II.
- C: Se confunde el nivel de significación con la probabilidad de alcanzar un resultado correcto. Esto puede estar asociado a la idea de asignar probabilidades pequeñas a

sucesos no deseables. Un suceso de baja probabilidad no suele ocurrir y, por su incidencia, sería de poco interés. En un contraste de hipótesis, justamente que este suceso ocurra, a pesar de su baja probabilidad, es lo que justifica el rechazo de la hipótesis.

D: Si se confunden las dos probabilidades condicionales $P(\text{rechazar } H_0 / H_0 \text{ cierta})$ y $P(H_0 \text{ cierta} / \text{rechazada } H_0)$.

Item 6: Un científico elige usar siempre el 0.05 como nivel de significación en sus experimentos en inferencia estadística. Esto significa que a la larga:

- A: El 5% de las veces, rechazará la hipótesis nula.
- B: El 5% de veces que rechace una hipótesis nula no se habrá equivocado.
- C: Habrá rechazado indebidamente la hipótesis nula sólo el 5% de las veces.
- D: Habrá aceptado una hipótesis nula falsa el 95% de las veces.

Se plantea en este ítem el nivel de significación como probabilidad de rechazar una hipótesis nula siendo cierta, en un contexto de repetición de un gran número de contrastes y de las expectativas en la replicabilidad de los resultados. Según Kahneman y cols. (1982), se suele confiar en exceso en la replicabilidad de los resultados experimentales, y ello es debido a la heurística de la representatividad y, más concretamente, a la creencia en la “ley de los pequeños números”. Los autores muestran cómo estos sujetos esperan una reproducibilidad de las características de la población en las muestras, independientemente del tamaño que tengan estas muestras. Por ello esperan confirmar el resultado de un contraste de hipótesis cuando se vuelve a repetir el contraste tomando muestras adicionales de la misma población. (Vallecillos, 1996)

Los distractores presentados en este ítem son:

A: Confusión de α con la probabilidad de rechazar la hipótesis nula, sea o no falsa. Es decir no se tiene en cuenta la condicionalidad planteada en la definición del nivel de significación, α .

B: Confusión de los condicionales.

D: Confusión entre los sucesos “rechazar una hipótesis nula cierta” y “aceptar una hipótesis nula falsa”.

Las respuestas de este ítem se compararon con las de los ítems 9 y 11 para analizar su consistencia.

- **Determinación del nivel de significación:**

Item 5: El nivel de significación en un contraste de hipótesis no puede ser justificado teóricamente desde la estadística, sino que es una libre elección de quien aplica el contraste *V/F*.

Con este ítem de Verdadero / Falso, se quiere determinar hasta qué punto los alumnos son conscientes de que en el proceso de inferencia ciertas decisiones deben ser tomadas por el experimentador en función del conocimiento e interés del problema de su disciplina y, que además, la estadística no puede proporcionar reglas para ello. El uso frecuente de niveles de significación .05 y .01 es cuestión de convenio y no se justifica por la teoría matemática. Si consideramos el contraste de hipótesis como proceso de decisión (la visión de N-P), debemos especificar el nivel de significación antes de llevar a cabo el experimento y esta elección determina el tamaño de las regiones críticas y de aceptación que llevan a la decisión de rechazar o no la hipótesis nula. En cambio, Fisher (1956) sugirió publicar el valor p exacto obtenido en cada experimento particular, lo que, de hecho, implica establecer el nivel de significación después de llevar a cabo el experimento.

- **Definición del nivel de significación:**

Los siguientes ítems 9 y 11 del cuestionario fueron dispuestos para poder analizar este concepto.

Item 9: 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)

Con este ítem se trata de analizar la confusión entre los condicionales en la definición del nivel de significación y, junto con el siguiente, poder comprobar la interpretación de esta definición.

Item 11: 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* (Razona la respuesta).

Aquí se enuncia la definición del nivel de significación como probabilidad de error de tipo I. Estos dos ítems tienen una redacción muy parecida, presentando el último una interpretación frecuencial del nivel de significación, mientras que en el anterior se han intercambiado los dos sucesos que definen la probabilidad condicional.

Resultados y Discusión

Los cinco ítems anteriores, como ya se ha aclarado, forman parte de un cuestionario más amplio utilizado en una muestra piloto de 125 estudiantes, pertenecientes a las carreras de Ingeniería Agronómica, Licenciatura en Biología y de Microbiología, y a 15 estudiantes del doctorado en Cs. Biológicas, de la UNRC. La Tabla 1 muestra los resultados obtenidos. El análisis de los mismos y de los distractores elegidos por los alumnos en los distintos ítems permite identificar las siguientes dificultades más frecuentes reveladas por los dos tipos de estudiantes:

- Interpretación del nivel de significación.

- Definición del nivel de significación.

La interpretación errónea más extendida consiste en intercambiar los dos términos de la probabilidad condicional. Para el ítem 3, la opción D es erróneamente la más elegida. Para el ítem 9, lo fue interpretar el nivel de significación como la probabilidad de que la hipótesis nula sea cierta si hemos tomado la decisión de rechazarla.

La semejanza en la redacción de los ítems 9 y 11, uno verdadero y el otro no, amerita el análisis conjunto de ambos, ya que permite comprender mejor la confusión de los condicionales que se presenta en la definición de nivel de significación. Cabe señalar que un alto porcentaje de los encuestados manifestaron no identificar las diferencias existentes entre estos dos ítems y sólo un 32% de los alumnos de grado y un 40% de los de postgrado respondieron correctamente ambos ítems.

Tabla 1. Porcentaje de resultados para cada opción, según ítem y tipo de alumno.

ITEM	A		B		c		d		No Contes		VERD		FALSO	
	Gra	Post	Gra	Post	Gra	Post	Gra	Post	Gra	Post	Gra	Post	Gra	Post
3	26	40	3	7	6	2	65	51	---	-	--	---	--	----
5	----	----	---	----	----	----	----	----	6	10	67	80	2	10
6	8	----	28	33	50	40	7	7	7	20	----		--	---
9	----	----	---	----	----	----	----	----	8	40	60	35	3	40
11	----	----	---	----	----	----	----	----	10	40	49	47	4	25

Nota: En negrita % respuestas correctas

Conclusiones

Las dificultades epistemológicas de los contrastes de hipótesis se ponen de manifiesto con las controversias entre las distintas escuelas que, por lo general, no son tenidas en cuenta en el proceso de su enseñanza. Esta enseñanza suele consistir en una mezcla poco armonizada de estos enfoques. Por otra parte, se pudo observar que aunque α es una probabilidad condicional bien definida, la expresión “Error Tipo I” no está redactada como una condicional, ni indica a cuál de las dos combinaciones posibles de los sucesos que intervienen se refiere. En consecuencia, cuando se rechaza H_0 el concepto “Error de Tipo I” viene a la mente en forma inmediata pero la distinción entre las dos direcciones opuestas de la probabilidad condicional se olvida. Esto lleva a interpretar el nivel de significación como la probabilidad de la conjunción de dos sucesos “la hipótesis nula es cierta” y “la hipótesis nula es rechazada”. De acuerdo a los resultados de esta investigación, se considera conveniente que al momento de definir el nivel de significación, se subraye que el nivel mismo es una probabilidad condicional y que luego, al interpretar los resultados en una determinada situación, se especifiquen los errores que surjan al rechazar, o no, H_0 como probabilidades condicionales y no, simplemente, como $P(\text{Error tipo I})$ o $P(\text{Error tipo II})$.

Reconocimientos

Este trabajo se realizó dentro del marco de un proyecto subsidiado por UNRC y del Proyecto de tesis del Doctorado en Ciencias que se está realizando en CICATA.

Bibliografía

- Batanero, C. (2000). Traducción del artículo: Controversies around the role of statistical tests in experimental researcher. *Mathematical Thinking and Learning*. 2000, 2(1-2), 75-98.
- Christensen, R. (2005). Testing Fisher, Neyman, Pearson, and Bayes. *American Statistical Association*. Vol.59, N°2
- Falk, R. & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology*, 5 (1), 75-98.
- Gómez Villegas, M.A. (2005). *Inferencia Estadística*. Ediciones Díaz de Santos.
- Ito, P. K. (1999). *Reaction to invited papers on statistical education and the significance tests controversy*. Ponencia invitada en la Fifty-Second International Statistical Institute Session, Helsinki, Finland.
- Kahneman, D., Slovic, P., y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Lêcoutre B. (2006). Training Students and Researchers in Bayesian Methods. *Journal of Data Science*, 4, 207-232.
- Moore, D.S. (1998a). *Estadística Aplicada Básica*. Barcelona. Antoni Bosch editor.
- Moore, D.S. (1998b). Incertidumbre. En L.A. Steen (Ed.) *La enseñanza agradable de las matemáticas*. Limusa. México
- Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Granada: Comares.