Significados del contraste de hipótesis, configuraciones epistémicas asociadas y algunos conflictos semióticos

Meanings of hypothesis testing, linked epistemic configurations and some semiotic conflicts

Carmen Batanero[1], Carmen Díaz[2] y Mª del Mar López-Martín[1]

[1]Universidad de Granada, [2]Universidad de Huelva

Resumen

Aunque la inferencia estadística es una herramienta esencial en muchas ramas de la actividad humana, y su enseñanza es generalizada, una extensa bibliografía crítica su uso inadecuado. Ello ha originado una línea de investigación y desarrollo denominada *inferencia informal* que está cobrando gran importancia. En este trabajo se resumen tres de las aproximaciones actuales al contraste de hipótesis y se utilizan algunas ideas del enfoque ontosemiótico con un doble objetivo: a) Identificar algunos componentes de las configuraciones epistémicas ligadas a dichos significados del contraste de hipótesis; b) Explicar mediante la idea de conflicto semiótico algunos errores y sesgos de razonamiento que se han descrito en las investigaciones sobre el tema. Se finaliza con algunas recomendaciones para la enseñanza del contraste de hipótesis.

Palabras clave: Contraste de hipótesis; diversos significados institucionales; enfoque ontosemiótico, configuraciones epistémicas y cognitivas; conflicto semiótico; idoneidad didáctica.

Abstract

Although statistical inference is an essential tool in many human activities, and its teaching is generalised, a wide bibliography criticises its incorrect use. These facts let to a research and development focus in what is being called informal inference, which is taking a great relevance. In this work, we summarise three current approaches to statistical tests and use some ideas from the onto-semiotic approach with a double aim: a) Identifying some components in the epistemic configurations linked to these meanings of statistical tests; b) Using the idea of semiotic conflicts to explain some reasoning errors and biases described in previous research. We finish with some recommendations for the teaching of statistical tests.

Keywords: Statistical tests; different institutional meanings; onto-semiotic approach; epistemic and cognitive configurations; semiotic conflict's.

1. Introducción

La inferencia estadística se incluye en la actualidad en el segundo curso de bachillerato de Humanidades y Ciencias Sociales, la mayoría de las carreras universitarias y estudios de postgrado. Además, en las pruebas de acceso a la universidad de Matemáticas Aplicadas a las Ciencias Sociales se ha venido proponiendo con frecuencia un problema sobre contraste de hipótesis (López-Martín, Batanero, Díaz-Batanero y Gea, 2016).

En lo que sigue analizamos, en primer lugar, tres aproximaciones diferentes del contraste de hipótesis, incluyendo el enfoque didáctico dentro de la línea de *inferencia informal* (ver, por ejemplo, Batanero, 2013; 2015; Noll, Gebresenbet y Glover, 2016; Rossman, 2008; Rubin, Hammerman y Konold, 2006; Zieffler, Garfield, del Mas y Reading, 2008). Seguidamente utilizamos la idea de conflicto semiótico para explicar algunos errores y sesgos de razonamiento que se han descrito en la investigación sobre

Batanero, C., Díaz, C. y López-Martín, M. M. (2017). Significados del contraste de hipótesis, configuraciones epistémicas asociadas y algunos conflictos semióticos. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M.M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos.* Disponibles en, enfoqueontosemiotico.ugr.es/civeos.html

el contraste de hipótesis o que hemos encontrado en algunas propuestas de la denominada inferencia informal. Finalmente se incluyen unas recomendaciones para la enseñanza del contraste de hipótesis.

2. Significado institucional de un objeto matemático y configuraciones epistémicas asociadas

Para llevar a cabo este análisis es importante partir de un marco teórico fuerte, que considere tanto el significado como la comprensión de un objeto matemático más allá de la simple definición del mismo. En nuestro caso partiremos del enfoque ontosemiótico, en el que hemos venido trabajando a lo largo de 20 años. Este enfoque surge con la finalidad de proporcionar apoyo a un programa de investigación sistemática en didáctica de la matemática, e incluye diversos instrumentos que permiten analizar los componentes de los procesos de enseñanza y aprendizaje. En este trabajo nos centramos en las ideas de objeto matemático, práctica y significado, que han sido analizadas desde los comienzos del desarrollo del marco teórico (por ejemplo, en Godino 2002; Godino y Batanero, 1994; 1998; Godino, Batanero y Roa 2005; Godino, Batanero y Font, 2007). Ya en Godino y Batanero (1999) tomábamos de Vygostki (1934) la idea de que el análisis de la actividad matemática debe partir del estudio previo del significado de las palabras que usamos en dicha actividad, pues estas reflejan la unión de pensamiento y lenguaje.

Este análisis parte de las ideas de problema y práctica; considerada ésta como toda actuación o expresión (verbal, gráfica, etc.) realizada por alguien para resolver problemas matemáticos, comunicar a otros la solución obtenida, validarla o generalizarla a otros contextos y problemas (Godino y Batanero, 1994, p. 334). Los objetos matemáticos emergen de las prácticas, que pueden ser personales o institucionales (compartidas en una institución) y el significado de un objeto se concibe como el sistema de prácticas asociadas al mismo (Godino, Batanero y Font, 2007). El término objeto se considera de una forma muy amplia, diferenciándose, en un primer nivel, los siguientes tipos, que se relacionan entre sí, como muestra la Figura 1[1]:

- *Los problemas.* Son aquellos de los que emerge un objeto. En el caso del contraste de hipótesis serían las situaciones en que se desea poner a prueba una conjetura o hipótesis en una ciencia de tipo empírico.
- *El lenguaje matemático.* Permite representar el objeto, dentro de éste bloque se incluiría las palabras como contraste, hipótesis, experimento, dato; símbolos y expresiones algebraicas como, por ejemplo, $N\left(\mu, \sigma/\sqrt{n}\right)$; lenguaje gráfico (representaciones gráficas de la distribución de datos, la distribución muestral o la distribución en la población, etc.).
- *Los procedimientos*. Algoritmos o estrategias que permiten operar con los datos para resolver el problema o generalizarlo. Por ejemplo, recoger datos de una muestra, calcular su media, calcular una probabilidad con las tablas de la distribución o con ordenadores.

[1] Véase las diferentes referencias citadas para una definición más precisa de objeto y significado, así como de las diferentes facetas desde las cuáles se puede diferenciar un objeto matemático o de los procesos matemáticos asociados. En este trabajo, nos restringimos a una definición elemental y simple de dichos términos.

- *Las definiciones del objeto, sus propiedades características y las relaciones con otros objetos*. Por ejemplo, población y muestra, variable aleatoria y estadística, estadístico muestral y parámetro poblacional o las relaciones y propiedades de los mismos.
- *Los argumentos.* Aquellos que usamos para probar las propiedades o relaciones o demostrar las soluciones de los problemas y, en general, para comunicar a otras personas nuestros resultados. En el trabajo de un estadístico se utilizaría un razonamiento de tipo deductivo para llevar a cabo el procedimiento de contraste o para demostrar algunas propiedades; en la enseñanza secundaria, se podría utilizar la simulación para mejorar la comprensión de los estudiantes.

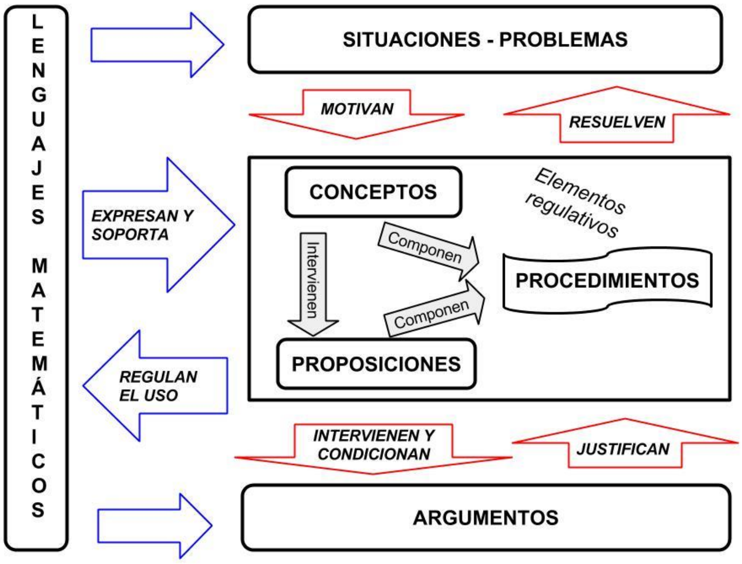

Figura 1. Configuración de objetos primarios (Godino, Batanero y Font, 2008)

La enumeración anterior, incluso sencilla, muestra el carácter polifacético de los objetos matemáticos, y la necesidad de tener en cuenta los tipos de objetos anteriormente descritos al organizar la enseñanza. Por supuesto, para cada uno de ellos son posibles diferentes niveles de formalización, dependiendo del nivel educativo. No sería lo mismo el trabajo con el contraste de hipótesis en bachillerato, la estadística profesional o la investigación en estadística. Cada una de las anteriores situaciones define una institución y hablamos del significado institucional del contraste de hipótesis. Por otro lado, en el marco teórico se diferencia entre el significado institucional y personal de un objeto matemático, para tener en cuenta la comprensión particular que una persona, dentro de la institución, puede alcanzar del objeto y que puede no coincidir con el significado institucional del mismo[2].

Por otro lado, la inferencia estadística (y el contraste de hipótesis que es parte de ella) tiene un estatuto particular, en el sentido que, a lo largo de la historia, incluso dentro de la actividad matemática profesional podemos describir diversas acepciones (que interpretamos en nuestro marco teórico como significados diferenciados). En lo que sigue, comenzamos describiendo brevemente la relación de la inferencia con el

[2] Además se tiene en cuenta diversos tipos de significado, tanto institucional (de referencia, implementado, pretendido, evaluado), como personal.

problema de la justificación de la inducción empírica y seguidamente analizamos algunas de estas acepciones y los objetos matemáticos que las caracterizan

3. Inferencia estadística y la justificación de la inducción empírica

La inferencia estadística se creó como respuesta al problema consistente en obtener un conocimiento general (una nueva teoría científica) a partir del análisis de casos particulares (inducción empírica). Esto es, nace con la búsqueda de métodos que permitan justificar el razonamiento inductivo y la extensión de sus conclusiones, problema de gran importancia en las ciencias empíricas (Cabriá, 1994; Mayo y Cox, 2006; Rivadulla, 1991). Dicho problema fue estudiado desde Aristóteles, aunque fue Hume el primero que los planteó con claridad en 1739, argumentando que ningún conjunto de observaciones pasadas puede justificar sucesos que ocurran en el futuro (Hacking, 2006).

En el razonamiento inductivo empírico, la conclusión se obtiene del análisis de casos particulares que se tratan de generalizar; en consecuencia, dicha conclusión es más amplia que las premisas, y puede ser falsa, aunque se parta de premisas ciertas. La justificación de la validez de estas conclusiones ha sido motivo de discusión en filosofía y estadística, sin que se haya obtenido una solución aceptada por consenso (Black, 1979; Cabria, 1994; de la Fuente y Díaz, 2004).

Supongamos, por ejemplo, que el responsable sanitario de un distrito quiere probar que el índice de masa corporal (IMC) de los adolescentes en su distrito es adecuado, es decir, no supera el valor fijado de 25. Sabe que la variable sigue una distribución normal con desviación típica igual a 5, ¿Cómo podría probar esta conjetura? Para resolver problemas similares a éste, Popper (1934/ 1967) propuso poner a prueba la hipótesis mediante experimentos u observaciones y comparar los datos obtenidos con los patrones deducidos de la hipótesis, para aceptarla, provisionalmente, si los datos siguiesen estos patrones. En el ejemplo dado, se comenzaría tomando datos del IMC en muestras sucesivas de adolescentes; si en todas estas muestras obtenemos un IMC medio dentro del rango esperado (menor que 25), se aceptaría la conjetura del director sanitario, provisionalmente. Pero es importante comprender que la confirmación de una teoría a partir de datos empíricos nunca es definitiva, porque los datos futuros podrían contradecirla. En cambio, si los datos del experimento se apartasen del patrón esperado, la teoría sería refutada, por lo que el rechazo de la hipótesis tiene mayor fuerza que su confirmación.

El comienzo del desarrollo de la inferencia estadística es anterior a Popper; entre 1764 y 1786, Bayes y Laplace resolvieron independientemente lo que hoy conocemos como problema de la probabilidad inversa (obtención de la distribución a posteriori de un parámetro a partir de datos de una muestra y de la distribución de probabilidad a priori del parámetro), que corresponde a la metodología bayesiana (Hald, 2008). Laplace también desarrolló el primer test de significación para la media, para el caso particular de una distribución uniforme de datos. Posteriormente, entre 1809 y 1929 Laplace y Gauss descubrieron que la distribución aproximada de las medias muestrales para muestras grandes es aproximadamente normal.

4. Configuraciones epistémicas ligadas a tres significados institucionales del contraste de hipótesis

Las ideas de Popper tuvieron una gran influencia en el desarrollo de la inferencia frecuencial, sistematizada entre 1920 y 1940 por Fisher, Neyman y Pearson, quienes desarrollan su teoría sobre los contrastes de hipótesis (Batanero y Borovcnik, 2016). Dado que mediante un razonamiento inductivo no es posible llegar a la certidumbre de una proposición (verdad cierta), estos y otros autores intentaron calcular la probabilidad de que una hipótesis sea cierta (verdad probable) (Batanero, 2000; Rivadulla, 1991). Ahora bien, diferentes escuelas de estadística han proporcionado diferentes soluciones al problema planteado. En las siguientes secciones describimos las metodologías frecuencial y Bayesiana.

Dentro de la inferencia frecuencial hay dos concepciones sobre los contrastes estadísticos: a) Las pruebas de significación, que fueron introducidas por Fisher y b) Los contrastes como reglas de decisión entre dos hipótesis, que fue la concepción de Neyman y Pearson. Estas aproximaciones no se diferencian en lo que concierne a los cálculos, pero si en sus objetivos.

La probabilidad de una hipótesis no tiene sentido en la inferencia frecuencial, porque en ella la probabilidad se interpreta como el límite de la frecuencia relativa en un gran número de repeticiones independientes del experimento aleatorio. Pero una hipótesis es siempre cierta o falsa. En el ejemplo dado, el IMC medio en la población superará o no el valor 25, no teniendo sentido concluir que la hipótesis es cierta un porcentaje de veces.

4.1. Metodología de Fisher: El test de significación como refutación empírica de una hipótesis

El *test de significación* fue propuesto por Fisher en su libro “The design of experiments, publicado en 1935, como procedimiento que permite rechazar una hipótesis, con un cierto *nivel de significación.* Fisher (1935/1971) introduce su teoría de las pruebas de significación, que resumimos en lo que sigue.

Supongamos que se quiere comprobar si una cierta hipótesis H_0 (hipótesis nula) es cierta. Se suele tomar como hipótesis nula, o de no efecto, la contraria de la que se pretende probar (en nuestro ejemplo, la hipótesis nula sería que el IMC medio no supera el valor 25). Generalmente la hipótesis se refiere al valor supuesto de un *parámetro,* pero no se tiene acceso a toda la *población,* sino sólo a una muestra de la misma. En el ejemplo, estamos interesados en el IMC medio, μ, en toda la población de adolescentes, pero solo podemos tomar una muestra; calcularemos la media $\bar{x}$ en la muestra y lo compararemos con el valor supuesto de la media en la población.

Para poner la hipótesis a prueba se organiza un experimento aleatorio asociado a H_0 y se considera un cierto suceso S que puede darse o no en este experimento y que es muy improbable de ser cierta la hipótesis. El experimento en el ejemplo sería recoger una muestra aleatoria de 225 adolescentes y calcular la media de su IMC. Se sabe que si H_0 fuese cierta (si el IMC medio en la población no supera el valor 25), hay muy poca probabilidad de que ocurra S (obtener un valor medio en la muestra mucho mayor que 25). En el ejemplo comparamos la media en la muestra con la supuesta media en la población, permitiendo así medir la discrepancia entre los datos muestrales y la hipótesis planteada.

Supongamos que, realizado el experimento, ocurre precisamente *S*, es decir, obtenemos, por ejemplo que $\bar{x} = 26$, que es muy improbable, bajo la suposición de que la hipótesis nula H_0 es cierta. En esta situación encontramos dos posibles conclusiones:

- La hipótesis $H_0 : \mu \leq 25$ era cierta y ha ocurrido *S* (obtener un valor medio en la muestra mucho mayor que 25), a pesar de su baja probabilidad, *p*=0.0013[3] (pues un suceso improbable no es imposible).
- La hipótesis H_0 era falsa.

Lo razonable sería rechazar la afirmación del director sanitario, pues si el IMC en su distrito fuese igual o menor a 25; sólo 13 de cada 10000 muestras de 225 adolescentes darían un índice medio igual o mayor que 26.

El razonamiento que apoya un test de significación parte de la suposición de que la hipótesis nula es cierta. En dicho caso, al variar los datos aleatoriamente, el estadístico define una distribución que recibe el nombre de *distribución muestral*, (Cabriá, 1994; Batanero, 2000). Un test de significación efectúa una división entre los posibles valores del estadístico en dos clases: resultados estadísticamente significativos (para los cuales se rechaza la hipótesis) y no estadísticamente significativos (para los cuáles no se puede rechazar la hipótesis) (Ridavulla, 1991).

En resumen, en el enfoque de Fisher: 1) El interés es rechazar la hipótesis nula; 2) No se identifica una hipótesis alternativa concreta (Batanero y Díaz, 2006); 3) No hay un criterio estándar sobre qué es un "suceso improbable"; 4) El valor de la probabilidad por debajo de la cual rechazamos la hipótesis (nivel de significación) lo fija el investigador según su juicio subjetivo y su experiencia (suele ser común tomar un nivel de significación α=0.05).

4.2. Metodología de Neyman y Pearson: contraste de hipótesis como regla de decisión

Neyman y Pearson por su parte estaban interesados en encontrar la regla óptima de decisión para elegir entre una hipótesis dada H_0 y otra hipótesis alternativa, denotada por H_1 (Rivadulla, 1991). Este enfoque tiene más sentido cuando se trata de una prueba que se repite muchas veces en las mismas condiciones. Supongamos, por ejemplo, un contexto de control de calidad: Estamos en un proceso de llenado de paquetes de azúcar de 1 kg, donde se supone una distribución normal N (0,1). Estamos interesados en diferenciar entre dos hipótesis:

> H_0: El proceso está controlado, es decir, los paquetes tienen un peso medio de 1kg ($\mu = 1$).
>
> H_1: Se ha descontrolado el proceso, es decir, los paquetes tienen un peso medio mayor o menor de 1 kg ($\mu \neq 1$).

Supongamos que, para realizar un adecuado control de calidad, todos los días se toma una muestra aleatoria de 30 paquetes con el fin de obtener el peso de cada uno (pues sería muy caro pesarlos todos) y cada día se repite el mismo procedimiento para ver si

---

[3] Al tipificar la media muestral en el ejemplo, y como la muestra es muy grande podemos usar la distribución normal N (0,1) para calcular la probabilidad de obtener un valor igual o mayor que 3 (valor obtenido en la tipificación, $(25-26)/(5/\sqrt{225})$). $P(Z \geq 3 \mid H_0) = 0.0013$.

paro o no la producción. Teniendo en cuenta el beneficio de la empresa, nos interesará encontrar la regla que a la larga, repitiendo el proceso muchas veces, dé el menor número de errores posibles.

Por ello contemplan dos posibles decisiones respecto a H_0: rechazar esta hipótesis, asumiendo que es falsa, aceptando la alternativa, o abstenerse de esa acción. Las dos situaciones pueden llevar a un error que implica un coste.

- *Error tipo I.* Rechazar una hipótesis nula cuando es cierta (en nuestro caso, parar el proceso de producción, cuando el proceso en realidad está controlado). El coste implicado sería el de revisión de la maquinaria innecesariamente. La probabilidad de cometer este tipo de error o *nivel de significación* α se fija antes de realizar el contraste y es una constante. Generalmente se trabaja con α=0.05.

- *Error tipo II.* Aceptar una hipótesis nula que de hecho es falsa (considerar que el proceso está controlado, cuando no lo está). Si supongo que el proceso está controlado cuando no lo está, puedo estar vendiendo más o menos peso del que cobro; aunque vender un menor peso podría parecer no grave, puedo perder mi calidad o mi imagen. La probabilidad de cometer este tipo de error se denomina Beta (β) y su complemento ($1 - \beta$) recibe el nombre de la *potencia* del contraste. Dicha probabilidad β es variable, ya que su valor depende del valor del parámetro que se está contrastando (generalmente desconocido). En el ejemplo que se está tratando, si en vez de envasar paquetes de 1 kg, el proceso fabrica paquetes de 999 gramos, la probabilidad de error tipo II es alta, porque hay poca diferencia entre el peso medio supuesto del paquete (1 kg) y la media producida (999 gramos). Sin embargo, si estamos fabricando paquetes de 1500 gramos será difícil suponer que el proceso está controlado, ya que el peso medio de los 30 paquetes será muy cercano a 1500 gramos y existirá diferencias con respecto al valor supuesto del parámetro.

4.3. Metodología Bayesiana

En la inferencia bayesiana la probabilidad de una hipótesis es un grado de creencia personal que oscila entre 0 (falsedad absoluta) y 1 (certeza absoluta) (Gigerenzer, 1993; Lecoutre, Lecoutre y Poitevineau, 2001). Por ello puede ser distinta para diferentes personas, pero, además, se puede revisar su valor a partir de los datos.

Una diferencia con la inferencia frecuencial es que el parámetro (en el ejemplo, el IMC medio de la población de adolescentes) es una variable aleatoria a la cual se la asigna una distribución de probabilidad. De hecho se consideran dos tipos de distribución del parámetro:

- *Distribución a priori o inicial.* Representa el grado de creencia inicial en los diferentes valores del parámetro antes de recoger datos de experimentos. Si no se tiene ninguna creencia previa se suele emplear como distribución inicial la distribución uniforme.

- *Distribución a posteriori o final.* Representa el grado de creencia en la hipótesis una vez se han recogido los datos. El teorema de Bayes servirá para combinar la probabilidad inicial con los datos y llegar a la distribución a posteriori. Este teorema se puede representar simplificadamente en la forma siguiente (Díaz, 2005), Distribución final = K x Verosimilitud x Distribución inicial,

donde K es una constante definida como, $K = \frac{1}{\sum_i P(A_i)\cdot P\left(B\,|\,A_i\right)}$ (inversa del teorema de la probabilidad total) y la verosimilitud es la probabilidad de obtener un valor determinado de $\bar{x}$ (media muestral) para cada posible valor de μ (media en la población), es decir, la verosimilitud es $P(\bar{x}\,|\,\mu)$. Dado que la distribución final depende de la distribución inicial, el teorema de Bayes corrige progresivamente las diferentes asignaciones de la distribución inicial cuando aumentamos el tamaño de la muestra, de forma que se tiende a una distribución final no muy diferente aunque las distribuciones iniciales varíen.

En el ejemplo del IMC, se tiene:

- Puesto que no tenemos creencia previa sobre la distribución inicial del parámetro, suponemos que es una distribución uniforme (o rectangular), es decir, se considera que la distribución inicial es *K'*. Sabemos del enunciado que la desviación típica de la población es conocida y vale 5.
- Si se extrae de la población una muestra aleatoria de n elementos y su media fue $\bar{x}$, la verosimilitud de obtener esta media, se obtiene a partir de la distribución normal $N\left(\bar{x},\sigma/\sqrt{n}\right) = N\left(26,5/\sqrt{225}\right) = N\left(26,0.33\right)$. Por tanto, la distribución final o a posteriori se obtiene como:

Distribución final = $K \times N\left(\bar{x},0.33\right) \times K' = K'' \times N\left(\bar{x},0.33\right)$, siendo $K'' = K \times K'$.

En conclusión se puede decir que, en la metodología bayesiana, para contrastar la hipótesis nula $\mu \le 25$, lo que se hace es calcular la probabilidad de la hipótesis contraria mediante el uso de la distribución final del parámetro, es decir,

$$P(\mu > 25) = P\left(\frac{\mu - 26}{0.33} > \frac{25-26}{0.33}\right) = P(Z > -3) = 1 - 0.00135 = 0.99865$$

Como consecuencia, de los cálculos realizados se tiene que la probabilidad de que los adolescentes no tengan sobrepeso es $P(\mu \le 25) = 0.00135$. Puesto que es más probable que los adolescentes tengan sobrepeso que no lo tengan, se rechaza la hipótesis del responsable sanitario.

Nótese que hemos obtenido el mismo valor que al calcular el *p-valor* en la metodología frecuencial; esto ocurre porque hemos usado una distribución inicial uniforme (en otro caso no coincidiría). Pero la interpretación que tiene ahora esta probabilidad es muy diferente ya que indica el grado de creencia en la hipótesis. Por tanto en la metodología bayesiana podemos calcular la probabilidad de que una hipótesis sea cierta (entendida como nuestro gado de creencia en dicha hipótesis).

4.4. Una aproximación informal al contraste de hipótesis

Como vemos, el contraste de hipótesis, en cualquiera de las metodologías descritas requiere el uso de muchos conceptos y procedimientos, que dependen del problema[4] y la metodología seguida. Cuando un alumno de bachillerato o primer curso de universidad resuelve un contraste de hipótesis, con frecuencia no comprende todos los pasos, sino que los aplica en forma mecánica. Para paliar este problema y mejorar la comprensión son varias las investigaciones que proponen una aproximación informal a la inferencia (ver, por ejemplo, Batanero, 2013, 2015; Noll, Gebresenbet y Glover, 2016; Rossman, 2008; Rubin, Hammerman y Konold, 2006; Zieffler, Garfield, delMas y Reading, 2008).

Una de las posibilidades es utilizar un simulador para obtener una distribución muestral empírica, en lugar de trabajar con la distribución muestral teórica. El resto del procedimiento sería similar. Por ejemplo, si seguimos la metodología de Fisher y queremos contrastar la hipótesis de que el IMC medio en la población no supera el valor 25 (siendo la desviación típica 5) con ayuda del simulador, el alumno también debe recordar que, para realizar un contraste de hipótesis sobre la media de la población μ, ha de utilizar la media de la muestra; en este caso $\overline{x} = 26$. Pero no ha de realizar cálculo formal de probabilidad; por tanto el problema se simplifica bastante ya que no necesita recordar la fórmula de la desviación típica de la media, la fórmula de tipificación o lectura de la tabla de la distribución normal.

En el simulador únicamente tiene que introducir el dato supuesto de la media de la población, μ=25, y de la desviación típica, σ=5. A partir de dicha información se ha realizado la simulación de 1000 muestras cada una de 225 sujetos de la población normal N (25, 5) con Fathom (podríamos haber usado Excel u otro programa). En la Figura 2 hemos representado en un gráfico de puntos las 1000 medias muestrales. Esta distribución sería una distribución muestral empírica que tendería a la distribución muestral teórica cuando se aumentase mucho el número de simulaciones. Pero 1000 muestras nos da una distribución aproximada, con la cual podemos obtener una estimación sobre la probabilidad de obtener este suceso, si la hipótesis nula fuese cierta. De las 1000 muestras simuladas se comprueba que dos de ellas están por encima de 26, luego dicha probabilidad (*p-valor*) será:

$$p\text{-}valor = P(\overline{x} \geq 26) \mid H_0 \text{ cierta}) \cong \frac{2}{1000} = 0.002$$

Observamos una ligera diferencia de esta estimación de la probabilidad con la probabilidad exacta, calculada la sección 4.1. Ello es debido a que, en vez de utilizar la distribución exacta de la media muestral, estamos simulándola; por tanto, introducimos un pequeño error en el cálculo. Sin embargo, si el número de simulaciones, como en el ejemplo, es alto, los valores exactos y estimados de la probabilidad son muy parecidos.

En consecuencia, la simulación (en vez del cálculo formal) puede utilizarse al comenzar la enseñanza del contraste de hipótesis para poder concentrar al alumno en el aprendizaje de la lógica del proceso y en los conceptos que son necesarios: muestra y población, media muestral y poblacional, distribución de la media muestral, hipótesis nula, valor *p*. Es importante hacer ver al alumno que el *p-valor* viene dado por una probabilidad condicional ya que, en su estimación, se supone cierta la hipótesis nula. Esto es sencillo, pues utilizando el mismo simulador podemos cambiar la hipótesis de partida; por ejemplo, suponer que la media de la población μ=25.5 y el alumno

[4] Encontraremos, además, diferencias, según se trate de un contraste unilateral o bilateral y según el parámetro contrastado.

observaría como cambia la probabilidad, que seguiría siendo pequeña, pero no tanto como antes. Si la media de la población fuese 25.5 no sería ya tan raro obtener un IMC en la muestra igual a 26; en este caso es posible que no se pueda rechazar la hipótesis nula para un nivel de significación α=0.05

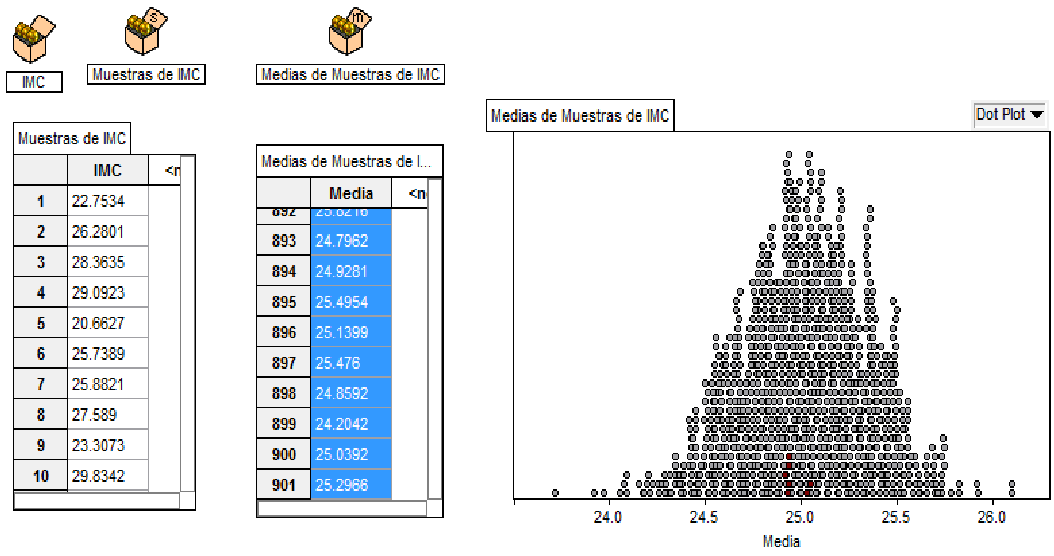

Figura 1. Simulación, con Fathom, de la distribución muestral el IMC (1000 muestras)

4.5. Síntesis

Tabla 1. Algunos objetos matemáticos ligados a los diferentes significados del contraste de hipótesis

Significados	Problema	Conceptos/Propiedades	Procedimientos
Test de significación	-Recabar evidencia en contra de una hipótesis	-Interpretación frecuencial de la probabilidad. -Hipótesis nula. -Distribución muestral. -El parámetro es fijo y desconocido.	-Fijar una hipótesis nula. -Calcular *p*-valor: probabilidad de obtener el valor dado u otro más extremo, bajo la hipótesis nula, usando la distribución muestral.
Contraste de hipótesis	-Encontrar una regla de decisión óptima entre dos hipótesis	-Interpretación frecuencial de la probabilidad. -Hipótesis nula y alternativa -Distribución muestral. -Error tipo I y II. -Potencia.	-Fijar una hipótesis alternativa. -Determinar las regiones crítica y de aceptación. -Cálculo de la potencia. -Determinar en qué región cae el estadístico.
Inferencia Bayesiana	-Actualizar la distribución del parámetro. -Calcular la probabilidad de una hipótesis.	-Interpretación subjetiva de la probabilidad. -Distribución a priori y a posteriori. -El parámetro es una variable aleatoria -Múltiples hipótesis.	-Asignar distribución a priori. -Teorema de Bayes. -Teorema Probabilidad Total. -Realizar cálculos con la distribución a posteriori. -Decidir qué hipótesis es más probable usando la distribución a posteriori.
Aproximación informal[5]	-Estimar la probabilidad de	-Interpretación frecuencial de la probabilidad.	-Obtención de muestras a partir de un simulador

[5] Se puede realizar una aproximación informal de cualquiera de las metodologías. En la tabla nos limitamos al test de significación de Fisher donde se sustituye la distribución muestral teórica por una

	que un modelo[6] produzca unos datos	-Modelo. -Distribución muestral empírica.	-Obtención de la distribución muestral empírica por simulación bajo el modelo supuesto. -Calcular p-valor, usando la distribución muestral empírica

En la exposición realizada, incluso resumida, se observa la complejidad del contraste de hipótesis, que implica la comprensión de muchos objetos matemáticos. Algunos como población y muestra, estadístico y parámetro, o hipótesis son comunes a todas las metodologías. Finalmente, cada aproximación cuenta con sus objetos matemáticos propios, por lo que podemos considerar los diferentes significados del contraste de hipótesis como configuraciones epistémicas diferenciadas. La Tabla 1 muestra algunos ejemplos de estos objetos.

5. Algunos ejemplos de conflictos semióticos relacionados con el contraste de hipótesis

La complejidad descrita explica los numerosos errores e interpretaciones incorrectas del contraste de hipótesis, que se han denunciado incluso en los trabajos de investigación (Batanero, 2000; Falk y Greenbaum, 1995; Harradine, Batanero y Rossman, 2011).

Otro componente del marco ontosemiótico es la idea de función semiótica, como correspondencia entre un antecedente (expresión, significador) y un consecuente (contenido, significado) establecido por un sujeto (persona o institución). Cualquier tipo de objeto matemático puede actuar como antecedente o consecuente en la función semiótica. Godino (2003) interpreta la comprensión en términos de las funciones semióticas que un sujeto puede establecer a propósito de un objeto dado. En Godino, Batanero y Font (2007) se describe el conflicto semiótico como cualquier disparidad de interpretación en el significado que dos sujetos asignan a la misma expresión. En lo que sigue interpretamos en términos de conflicto semiótico algunos errores sobre, el contraste de hipótesis descritos en la literatura y otros que, inconscientemente, se introducen en la enseñanza cuando se realiza con una aproximación informal.

5.1. Conflictos de aprendizaje

Un concepto que se suele comprender erróneamente es el nivel de significación α. Como se ha indicado anteriormente, es la probabilidad de rechazar la hipótesis nula, supuesta cierta H_0. En forma simbólica: $\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ es cierta})$. Supongamos que trabajamos con un valor α=0.05 o del 5%; esto quiere decir que si H_0 es cierta, la rechazamos 5 de cada 100 veces, o lo que es lo mismo en 100 días que hagamos el control de calidad y en los cuáles el proceso esté controlado, pararemos el proceso innecesariamente 5 días.

empírica obtenida por simulación. Además los conceptos no se explicitan, tan sólo se usan en forma intuitiva.

[6] Para evitar introducir términos técnicos como hipótesis en las aproximaciones informales, con frecuencia se habla de modelo, entendido coloquialmente, como el modelo implementado en el simulador.

Respecto a este concepto se pueden encontrar numerosos conflictos semióticos, es decir, errores de interpretación. El más frecuente consiste en intercambiar los dos términos de la probabilidad condicionada en la expresión anterior, es decir, interpretar α como la probabilidad de que la hipótesis nula sea cierta, una vez que la decisión de rechazarla se ha tomado, esto es, suponer que $\alpha = P(H_0 \text{ es cierta} \mid \text{se ha rehazado } H_0)$. Este tipo de error se ha encontrado en investigaciones llevadas a cabo con estudiantes (Birnbaum, 1982; Falk, 1986; Vallecillos, 1994) y en profesores de universidad responsables de la enseñanza de métodos de investigación (Krauss y Wassner, 2002). Subyace en este conflicto otro consistente en asignar a la probabilidad condicional la propiedad conmutativa (se pueden intercambiar sus términos).

Un segundo conflicto es la identificación de las funciones que en el contraste tiene la hipótesis nula y la hipótesis alternativa. Es decir, algunos estudiantes piensan que la hipótesis nula es la que queremos demostrar (no la que queremos rechazar).

Vallecillos (1999) describió las siguientes creencias que tenían sus estudiantes sobre el tipo de prueba que proporciona el contraste de hipótesis, cada una de las cuáles implica un conflicto semiótico:

- *El contraste de hipótesis en la metodología frecuencial es una prueba probabilística de la hipótesis.* Permite calcular la probabilidad de que una hipótesis sea cierta. Sin embargo, como hemos dicho, la probabilidad de una hipótesis no tiene sentido en inferencia frecuencial. Estos estudiantes darían una interpretación bayesiana a un procedimiento frecuencial.
- *El contraste estadístico es un método matemático; como tal, y al ser la matemática una ciencia exacta, al finalizar hemos probado la verdad o falsedad de una hipótesis.* Esta creencia, siempre errónea, supone un conflicto entre una demostración deductiva y el contraste de hipótesis.

5.2. Conflictos en la enseñanza

Algunos de los conflictos, presentados en la Sección 5.1, se transmiten involuntariamente en la enseñanza, en particular, algunas propuestas tratan de desarrollar el razonamiento inferencial informal con poco rigor. Algunos ejemplos de estos conflictos son los siguientes[7]:

- *Confusión entre frecuencia relativa y probabilidad.* Es frecuente encontrar en la literatura las expresiones probabilidad empírica o probabilidad frecuencial para referirse a la estimación frecuencial de la probabilidad a partir de la frecuencia relativa. Este conflicto ya fue denunciado por Chaput, Girard y Henry (2011).
- De la misma forma se encuentra en la literatura expresiones como *probabilidad teórica* para contraponerla con la frecuencial o subjetiva, mientras que toda probabilidad es teórica; lo empírico sería su estimación por medio de la frecuencia. Es igualmente desafortunado el término *probabilidad experimental*, puesto que todas las probabilidades son teóricas.
- *Confusión entre la distribución muestral teórica y empírica.* Es importante que el profesor comprenda los límites y no sólo las posibilidades de la simulación.

[7] Citaremos los conflictos encontrados en la literatura, sin incluir la referencia a los autores, que posiblemente no sean conscientes de los mismos.

La simulación introduce un error de estimación añadido a los errores tradicionales en el estudio del contraste de hipótesis (errores tipo I y II). Además, sólo podemos controlar este error, a medida que aumenta considerablemente el número de simulaciones, pues la distribución muestral empírica, obtenida mediante simulación, converge a la teórica cuando el número de muestras tomadas tiende a infinito. Sin embargo, para un número finito de simulaciones, las distribuciones muestrales empíricas varían de una simulación a otra. Luego, si cada alumno simula su propia distribución muestral, los resultados obtenidos por cada uno de ellos pueden ser diferentes, lo que puede añadir complejidad a la clase. Una forma de solventar este problema sería que el profesor acumule los resultados de todas las simulaciones para que la clase trabaje con una única distribución muestral más precisa.

- *Confusión entre probabilidad e inferencia estadística*. En algunos trabajos se plantean preguntas de tipo probabilístico, se consideran tales preguntas como tareas propias de la inferencia estadística, aunque falten los elementos mínimos que las caracterizan.
- En el caso del contraste de hipótesis, se simplifica excesivamente, sin considerar los posibles errores asociados, y se puede transmitir una visión muy determinista del contraste de hipótesis.
- En ocasiones no se diferencia entre la distribución estadística de datos en la muestra y la distribución de probabilidad en la población (detrás de las cuales hay dos objetos matemáticos diferenciados, variable estadística y variable aleatoria), sino se considera que se trata de dos puntos de vista diferente sobre la misma distribución.
- Otras veces no se resalta la naturaleza condicional del valor *p*, es decir, no se relaciona la probabilidad adecuadamente con la hipótesis supuesta y se produce de nuevo el conflicto consistente en intercambiar los términos en la probabilidad condicional al trabajar con el *p* valor y el nivel de significación. Al no explicitar los conceptos subyacentes, estos quedan sin fijarse en la mente de los alumnos.
- En otras propuestas de "inferencia informal", simplemente se pide tomar una decisión subjetiva sobre el rechazo de una hipótesis utilizando sólo procedimientos descriptivos; lo cual no sería una verdadera inferencia estadística. En esta situación encontramos un conflicto entre la inferencia y análisis exploratorio de datos.

Finalmente en muchos casos, se trata de sustituir todo el razonamiento probabilístico mediante la tecnología, sin enfatizar en los procesos aleatorios involucrados en la simulación. Tampoco esto sería una inferencia estadística, pues esta debe estar justificada por un modelo de probabilidad que ligue los datos a la población (Rossman, 2008). Por otro lado, en algunas propuestas se sustituye el aprendizaje de una de las metodologías de Fisher o Neyman-Pearson por el aprendizaje de la construcción de modelos de simulación en un software específico, añadiendo un tiempo innecesario al aprendizaje, pues, salvo problemas muy triviales, construidos específicamente para la enseñanza – y que no surgen de una investigación real- la construcción de dichos modelos no es tan simple. Si el estudiante construye un simulador inadecuado, la distribución muestral empírica obtenida no sería válida para realizar el contraste.

Un conflicto aún mayor, denunciado por Gigerenzer (1993) es la mezcla de elementos de las metodologías de Fisher, Neyman-Pearson y Bayesiana en la enseñanza, incluso en la enseñanza formal. Dichas metodologías tienen diferentes fines y son apropiadas en distintas situaciones que el profesor debe hacer ver a sus estudiantes, para promover un uso razonable de la inferencia estadística.

6. Conclusiones para la enseñanza e investigación

Los elementos utilizados del enfoque ontosemiótico nos han ayudado, por un lado, a analizar las diferencias fundamentales entre tres metodologías actuales de inferencia y una aproximación informal a una de ellas. Es claro que es posible una aproximación informal, pero rigurosa a cualquiera de las tres metodologías descritas. Sin embargo, a la hora de realizar esa aproximación es necesario, en primer lugar, seleccionar la metodología que se desea estudiar y presentarla con los elementos que la caracterizan. En segundo lugar hay que realizar un análisis profundo de cuáles elementos se pueden simplificar, y hasta qué punto, para no distorsionar el procedimiento de contraste de hipótesis ni la interpretación de sus resultados.

Por otro lado, la idea de conflicto semiótico nos ha permitido interpretar algunos errores descritos en el aprendizaje y práctica del contraste de hipótesis y, además, nos ha permitido resaltar otros que se transmiten en forma inconsciente en la literatura sobre inferencia informal. Es claro que la enseñanza actual de la estadística no lleva a un aprendizaje correcto y los estudiantes se limitan a memorizar las reglas de aplicación del contraste. Sería importante apoyar más la enseñanza sobre la intuición, para lo cual contamos con la simulación, que permite concretizar el muestreo aleatorio, el concepto de distribución muestral y su estimación empírica.

Simultáneamente, es importante que el profesor enfatice los pasos requeridos en el procedimiento de contraste, discuta con los estudiantes la naturaleza de las probabilidades que calcula (condicionales), así como la interpretación correcta de las probabilidades de error. En concreto, es necesario asegurar una idoneidad epistémica razonable, que no siempre se alcanza en las propuestas de enseñanza de inferencia informal. Dicha idoneidad epistémica implica que se incluyan en la enseñanza una muestra representativa de los objetos matemáticos implicados, que, aunque no sea completa, no los distorsione introduciendo objetos o propiedades que no tienen contrapartida en el significado institucional del concepto.

Puesto que la comprensión del contraste de hipótesis conlleva la de una gran cantidad de conceptos relacionados, es necesario contar con tiempo suficiente para desarrollar una adecuada comprensión en los estudiantes. Comenzando en la educación secundaria con el muestreo, se podría trabajar en el Bachillerato la distribución muestral y estimación puntual y por intervalo, dejando el contrate de hipótesis para la universidad. Es importante, también continuar la investigación sobre el razonamiento inferencial de los estudiantes, para determinar la enseñanza óptima del tema en cada etapa educativa, asegurando una adecuada idoneidad cognitiva de la misma.

Reconocimiento: Proyecto EDU2016-74848-P y grupo FQM126 (Junta de Andalucía).

Referencias

Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, *2*(1-2), 75-98.

Batanero, C. (2013). Del análisis de datos a la inferencia: Reflexiones sobre la formación del razonamiento estadístico. *Cuadernos de Investigación y Formación en Educación Matemática*, *8*(11), 277-291.

Batanero, C**.** (2015). Aproximación informal al contraste de hipótesis. M. Contreras (Ed.), II *Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatori*a. Granada: SEIEM.

Batanero, C. y Borovcnick, M. (2016). Statistics and probability in high school. Rotterdam: Sense Publishers.

Batanero, C. y Díaz, C. (2006). Methodological and didactical controversies around statistical inference. *Actes du 36iémes Journées de la Societé Française de Statistique*. CD ROM. Paris: Societé Française de Statistique.

Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics*, *4*, 24–27.

Black, M. (1979). *Inducción y probabilidad*. Madrid: Cátedra.

Cabriá, S. (1994). *Filosofía de la estadística.* Valencia: Servicio de Publicaciones de la Universidad.

Chaput, B., Girard, J. C., y Henry, M. (2011). Frequentist approach: modelling and simulation in statistics and probability teaching. En C. Batanero, G. Burrill, y C. Reading (Eds.), *Teaching Statistics in school mathematics-Challenges for teaching and teacher education* (pp. 85-95). New York: Springer.

Díaz, C. (2005). *Apuntes sobre inferencia bayesiana*. Granada: La autora.

Falk, R. (1986) Misconceptions of statistical significance, *Journal of Structural Learning*, *9*, 83–96.

Falk, R., y Greenbaum, C. W. (1995) Significance tests die hard: The amazing persistence ofa probabilistic misconception, *Theory and Psychology*, 5 (1), 75-98.

Fisher, R. A. (1971). *The design of experiments*. Edinburgh: Oliver y Boyd (trabajo original publicado en 1935).

de la Fuente, E. I., y Díaz, C. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*, Volumen especial 2004, 161-167.

Gigerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. En G. Keren y C. Lewis (Eds.), *A handbook for data analysis in the behavioural sciences: Methodological issues* (pp. 311 – 339). Hillsdale, NJ: Erlbaum.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathematiques 22* (2 y 3), 237-284.

Godino, J. D. (2003). *Teoría de las funciones semióticas*. Un enfoque ontológico-semiótico de la cognición e instrucción matemática. Departamento de Didáctica de la Matemática. Universidad de Granada. Disponible en, http://www.ugr.es/local/jgodino/indice_tfs.htm

Godino, J. D., y Batanero, C. (1994). Significado personal e institucional de los objetos matemáticos. *Recherches en Didactiques des Mathématiques*, *14* (3), 325-355.

Godino, J. D. y Batanero, C. (1998) Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education. In A. Sierpinska, and J. Kilpatrick (Eds.) *Mathematics Education as a Research Domain: A Search for Identity* (177-195). Dordrecht: Kluwer.

Godino, J. D. y Batanero, C. (1999). The meanings of mathematical objects as análisis units for didactics of mathematics. In I. Schwank (Ed.) *Proceedings of the First Conference of the European Society for Research in Mathematics Education* (Vol. 2, 232244). Osnabrück: Forschungsinstitut für Mathematikdidaktik.

Godino, J.D., Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education *ZDM, The International Journal on Mathematics*

Education, 39(1-2) 127-135.

Godino, J. D., Batanero, C. y Font, V. (2008). Un enfoque ontosemiótico del conocimiento y la instrucción matemática. *Acta Scientiae. Revista de Ensino de Ciências e Matemática, 10* (2), 7-37.

Godino, J. D., Batanero, C., y Roa, R. (2005). An onto-semiotic analysis of combinatorial problems and the solving processes by university students. *Educational Studies in Mathematics*, *60* (1), 3-36.

Hacking, I. (2006). *The emergence of probability. A philosopfhical study of early ideas about probability, induction and statistical inference*. New York: Cambridge University Press.

Hald, A. (2008). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935.* New York: Springer.

Harradine, A., Batanero, C., y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill, y C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 235-246). Springer Netherlands.

Lecoutre, B., Lecoutre M. P., y Poitevineau J. (2001). Uses, abuses and misuses of significance tests in the scientific community: Won't the Bayesian choice be unavoidable? *International Statistical Review*, *69,* 399-418.

López-Martín, M.M., Batanero, C., Díaz-Batanero, C., y Gea, M.M. (2016). La inferencia estadística en las pruebas de acceso a la universidad en Andalucía. *Revista Paranaense de Educaçao Matemática*, *5*(8), 33-59.

Mayo, D. G., y Cox, D. R. (2006). Frequentist statistics as a theory of inductive inference. *IMS Lecture Notes-Monograph Series, 49,* 77-97.

Noll, J., Gebresenbet, M., y Glover, E. D. (2016). A modeling and simulation approach to informal inference: successes and challenges. En D. Ben-Zvi (Ed.), The teaching and learning of statistics (pp. 139-150). New York: Springer.

Popper, K. R. (1967). *La lógica de la investigación científica.* Madrid: Tecnos (Obra original escrita en 1934).

Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropos.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7*(2), 5-19.

Rubin, A., Hammerman, J. K. L., y Konold, C. (2006). Exploring informal inference with interactive visualization software. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Cape Town, South Africa: International Association for Statistics Education. Online: www.stat.auckland.ac.nz/~iase/publications

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Proceedings of the 52 session of the International Statistical Institute* (Vol.2, pp. 201–204). Helsinki: International Statistical Institute.

Vygotski, L. S. (1934). *Pensamiento y lenguaje*. Madrid: Visor.

Zieffler, A., Garfield, J. B., delMas, R., y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 5-19. Online: www.stat.auckland.ac.nz/serj/.