

¿CÓMO PUEDE EL MÉTODO BAYESIANO CONTRIBUIR A LA INVESTIGACIÓN EN PSICOLOGÍA Y EDUCACIÓN?

**Carmen Díaz,
Carmen Batanero**

Universidad de Granada, España

Resumen

Numerosos autores han analizado el debate sobre los contrastes de hipótesis, así como las recomendaciones de diferentes asociaciones profesionales sobre su uso. Asimismo se han publicado revisiones y análisis exhaustivos de las posibilidades y criterios de uso de otros métodos inferenciales, como el análisis de potencia, o intervalos de confianza. En este artículo se complementan dichas publicaciones, con un análisis de las principales razones que apoyan la viabilidad de la metodología bayesiana y la forma en que puede complementar los análisis frecuenciales dentro de la psicología y las ciencias de la educación.

Palabras clave: Metodología de investigación, inferencia bayesiana, estudio de viabilidad

Abstract

Numerous authors have analysed the debate on statistical tests as well as the recommendations from different professional associations about its use. In the same way many revisions and exhaustive analyses of other frequentist methods such as power analysis, or confidence intervals determination have been published. In this work we complement these studies with an analysis of the main reasons supporting the viability of Bayesian methods and the way they can complement the frequentist methodology in Psychology and Education.

Keywords: Research Methods, Bayesian inference, analysis of suitability

Recibido: 02/08/2006 **Aceptado:** 02/11/2006

Introducción

El debate sobre el uso del contraste de hipótesis y la interpretación incorrecta de resultados estadísticamente significativos es muy extenso (Morrison y Henkel, 1970; Falk y Greenbaum, 1995; Harlow, Mulaik y Steiger, 1997; Ares, 1999; Borges, San Luis, Sánchez, y Cañadas, 2001; de la Fuente y Díaz, 2004). Como consecuencia, se publicaron recomendaciones (Levin y Robinson, 1999; Wilkinson, 1999; Fidler, 2002) sobre las normas que debieran seguir los editores de revistas a la hora de aceptar los trabajos para publicación.

Además de estas recomendaciones, se han publicado revisiones y análisis exhaustivos sobre las posibilidades que brindan otros procedimientos estadísticos que podrían sustituir o complementar los contrastes de hipótesis (Valera, Sánchez, Marín y Velandrino, 1998; Valera, Sánchez y Marín, 2000; Frías, Pascual, y García 2000; Pascual, García y Frías 2000). El objetivo de este trabajo es completar los estudios anteriores, aportando argumentos a favor del uso de métodos bayesianos, tema que apenas ha recibido atención en nuestra comunidad, aunque su utilización es recomendada por eminentes psicólogos y educadores como Edwards, Lindman y Savage (1963), Rozeboom (1970), Pruzek (1997), Rindskopf (1997) y Lecoutre (2006).

En lo que sigue comenzamos resumiendo y comparando las características de la inferencia clásica y bayesiana. A continuación presentamos los siguientes argumentos que aconsejan considerar la metodología bayesiana como una alternativa adecuada: a) la inferencia bayesiana no contiene mayor subjetividad que la de otros métodos estadísticos; b) proporciona la información que realmente interesa a los investigadores y c) existe software estadístico disponible que facilita la aplicación de esta metodología. Finalmente razonaremos que los conceptos básicos de esta metodología son asequibles para los estudiantes de psicología y educación, si se realiza el necesario esfuerzo didáctico.

Inferencia Frecuencial e Inferencia Bayesiana

La problemática asociada a la inferencia estadística es de tipo filosófico y se relaciona con la naturaleza incierta del conocimiento empírico, es decir, con la dificultad de justificar el razonamiento inductivo y sus conclusiones (Rivadulla, 1991). Ya que no es posible demostrar la veracidad de una hipótesis mediante un razonamiento inductivo (Cabriá, 1994; Hager, 2000), las diversas escuelas de inferencia estadística han tratado de aportar una solución matemática al problema de encontrar un criterio racional para aceptar o rechazar hipótesis, asumiendo ciertos riesgos.

Fisher, con sus pruebas de significación, y Neyman y Pearson, con el contraste de hipótesis, crearon metodologías basadas en la concepción frecuencial de la probabilidad. Dichos autores concibieron los parámetros a estimar en una población (por ejemplo la media), como constantes desconocidas. Imaginando la repetición infinita del experimento consistente en tomar una muestra del mismo tamaño de la misma población, llegaron a deducir la distribución muestral (que describe el comportamiento de los posibles valores del estadístico en el conjunto de todas las muestras) y de ahí, también deductivamente, realizaron inferencias sobre los valores plausibles de los parámetros en la población.

Supongamos, por ejemplo, que se desea estimar la media m de una distribución normal en la que se conoce la desviación típica σ . Se puede deducir que la distribución muestral de la media \bar{x} es normal $N(m, \sigma/\sqrt{n})$ y a partir de ahí, obtener un intervalo de confianza para el parámetro μ , $(\bar{x} - Z_{1-\alpha/2} \sigma / \sqrt{n}; \bar{x} + Z_{1-\alpha/2} \sigma / \sqrt{n})$ siendo Z un percentil de la distribución normal tipificada. Los límites de este intervalo de confianza son aleatorios (pues cambian de muestra a muestra), pero al construir todos los intervalos posibles para la población dada, se puede asegurar que un porcentaje igual a α por ciento incluirán al verdadero valor del parámetro, aunque no se puede decir nada sobre si el intervalo calculado en la muestra tomada lo cubre o no.

La inferencia Bayesiana está basada en la aplicación sistemática del Teorema de Bayes, cuya publicación en 1763 desconcertó a los matemáticos de la época. Mientras las concepciones anteriores de la probabilidad, tanto clásica (cociente entre casos favorables y posibles), como frecuencial (límite al cual tiende la frecuencia relativa) suponen un valor objetivo de la misma, la posibilidad abierta por el teorema de revisar las probabilidades iniciales en función de la nueva información, les hace perder dicho carácter objetivo (Hacking, 1975; Cabriá, 1994). La visión subjetiva, amplia también las aplicaciones de la probabilidad, no siendo ya un requisito la repetición de una experiencia en las mismas condiciones. Gradualmente, se desarrolla la distinción entre probabilidad frecuencial, empíricamente accesible a través de la frecuencia y probabilidad epistémica o grado de creencia en la ocurrencia de un suceso en un experimento único (Rouanet, 1998), mientras se conforman dos escuelas de inferencia.

La inferencia bayesiana considera que un parámetro q es una variable aleatoria y le asocia a una distribución inicial de probabilidades $p(q)$, de carácter epistémico, que indica el conocimiento (o falta de conocimiento) sobre q antes de tomar los datos. Si se considera $y = (y_1, \dots, y_n)$ un conjunto de datos, cuya *función de verosimilitud* $p(y/q)$ depende del parámetro, entonces la distribución condicional de q dados los datos observados y viene dada por el Teorema de Bayes:

$$(1) \quad p(\theta/y) = \frac{p(y/\theta)p(\theta)}{p(y)}$$

En la expresión (1) $p(y) = \sum p(y/\theta)p(\theta)$, donde la suma se extiende a través de todo el rango admisible de valores de θ (Box y Tiao, 1992; Lee, 2004). La distribución final $p(\theta/y)$ recoge toda la información sobre θ una vez conocidos los datos. El teorema de Bayes podría aplicarse sucesivamente en nuevos experimentos, tomando como probabilidades iniciales del segundo experimento las probabilidades finales obtenidas en el primero y así sucesivamente. Se habla así de "proceso de aprendizaje" (Box y Tiao, 1992).

Por ejemplo, supongamos de nuevo que queremos estimar la media μ de una distribución normal, en la que se conoce la desviación típica σ y supongamos también que la distribución inicial de μ uniforme (este caso se conoce como distribución inicial no informativa, pues suponemos que cualquier valor de μ es equiprobable, o lo que es lo mismo, no tenemos información inicial sobre el parámetro). En este caso, aplicando el teorema de Bayes se demuestra que la *distribución final* de la media es normal $N(\bar{x}, s/\sqrt{n})$, siendo \bar{x} el valor obtenido de la media muestral (Berry, 1995). A partir de esta distribución final, el intervalo de credibilidad para la media μ viene dado por: $(\bar{x} - Z_{1-\alpha/2}\sigma/\sqrt{n}; \bar{x} + Z_{1-\alpha/2}\sigma/\sqrt{n})$, siendo Z un percentil de la distribución normal tipificada.

En este ejemplo (y en otros casos de distribución inicial no informativa), la distribución final $p(\theta/y)$ para un cierto parámetro coincide con la distribución muestral, aunque su interpretación es muy diferente. Mientras el área en un intervalo de la distribución muestral es la probabilidad de que el estadístico de una muestra aleatoria entre todas las posibles en la población esté en dicho intervalo, el área en un intervalo en la distribución final es la probabilidad de que el parámetro esté en el intervalo en la muestra particular obtenida (Rindskopf, 1997). Como señala Cabriá (1994, p. 216) "*las inferencias bayesianas contienen su propia medida interna de la situación relacionada únicamente con la situación actual, sin necesidad de acudir a repeticiones de esa situación mediante distribuciones en el muestreo*".

Por otro lado, cuando se dispone de información inicial, los resultados de la inferencia bayesiana ya no coinciden con los de la frecuencial. En el ejemplo, si se supone que la distribución inicial para la media de la población sigue una curva normal $N(\mu_i, \sigma_i)$ y se conoce la desviación típica σ_i de la distribución inicial de la media, la distribución final también sigue la distribución normal $N(\mu_f, \sigma_f)$, donde los valores de la media y desviación típica de la distribución final vienen dados por (2), siendo n es el tamaño de la muestra, \bar{x} y s la media y desviación típica de la muestra.

$$(2) \quad \mu_f = \frac{\frac{n\bar{x}}{s^2} + \frac{\mu_i}{\sigma_i^2}}{\frac{n}{s^2} + \frac{1}{\sigma_i^2}} \quad \sigma_f = \frac{1}{\sqrt{\frac{n}{s^2} + \frac{1}{\sigma_i^2}}}$$

El intervalo de credibilidad para un coeficiente de credibilidad a viene dado por: $(\bar{x} - Z_{1-\alpha/2} \sigma_f / \sqrt{n}; \bar{x} + Z_{1-\alpha/2} \sigma_f / \sqrt{n})$, siendo Z un percentil de la distribución normal tipificada. Mientras que \bar{x} es un estimador insesgado de mínima varianza de la media poblacional, μ_f es un estimador sesgado. El sesgo viene dado por la expresión $\mu_f = \frac{\sigma_i^2}{ns^2 + \sigma_i^2} (\bar{x} - \mu_i)$ (Bolstad, 2004), es decir, disminuye con la diferencia entre la media de la distribución inicial y la media de la muestra, con las varianzas y tamaño de muestra y con el valor de la desviación típica en la distribución inicial.

Por otro lado la varianza del estimador es menor que la de la media muestral puesto que $\sigma_f^2 = \frac{1}{\frac{n}{s^2} + \frac{1}{\sigma_i^2}} = \frac{s^2}{n} \times \frac{n/s^2}{\frac{n}{s^2} + \frac{1}{\sigma_i^2}}$; por tanto es menor que $\frac{s^2}{n}$ al estar multiplicado por una cantidad menor que 1.

La aplicación sistemática del teorema de Bayes constituye el método principal de la inferencia bayesiana, cuyo objetivo básico es actualizar la distribución inicial de los parámetros. La distribución final es la esencia de la estimación bayesiana. A la pregunta, una vez vistos los datos, ¿qué sabemos del parámetro?, la respuesta es la distribución final, puesto que esta distribución sintetiza toda la información sobre el parámetro, recogidos los datos y contiene todas las inferencias que puedan hacerse del mismo (O'Hagan y Forster, 2004). El estimador puntual óptimo del parámetro será la esperanza del mismo en la distribución final, puesto que minimiza el error cuadrático esperado (O'Hagan y Forster, 2004).

La distribución final también permitirá calcular probabilidades de que los parámetros se encuentren en intervalos de valores dados (intervalos de credibilidad), así como calcular probabilidades de que ciertas hipótesis sean verdaderas o falsas. La inferencia bayesiana básica sobre una hipótesis es calcular su probabilidad final, al contrario que en inferencia clásica, donde las hipótesis se aceptan o rechazan, lo cual no es una inferencia, sino una decisión (O'Hagan y Forster, 2004).

La distribución predictiva o marginal $p(y) = \int p(y/\theta)p(\theta)d\theta$ permite también realizar predicciones sobre futuros valores de y que tienen en cuenta la incertidumbre sobre el valor del parámetro θ así como la incertidumbre residual sobre y cuando se conoce θ (Lee, 2004). Este tipo de probabilidad no puede calcularse en inferencia clásica (Bolstad, 2004).

Sobre la Subjetividad en los Métodos Bayesianos

Una diferencia fundamental de la inferencia Bayesiana respecto a la clásica es el carácter subjetivo (y no frecuencial) de las probabilidades, ya que no se plantea el problema de muestreo repetido ni requiere del concepto de distribución muestral. Las probabilidades subjetivas pueden definirse para cualquier proposición, mientras que la probabilidad frecuencial se define sólo

para sucesos en un espacio muestral (O'Hagan y Forster, 2004). Asimismo, el método bayesiano hace uso de toda la información previa disponible, mientras que en inferencia clásica esta información no se tiene en cuenta.

Puesto que el investigador especifica la distribución inicial, el enfoque bayesiano tiene en cuenta la perspectiva del investigador, su conocimiento del problema. No hay una única forma de elegir la distribución inicial, que condiciona los resultados de la inferencia y este hecho ha originado fuertes críticas al método bayesiano. Ello puede llevar a que distintos investigadores obtengan diferentes resultados de los mismos datos, en función de su conocimiento o experiencia previa. Para hacer frente a ellas se ha sugerido usar distribuciones no informativas al comenzar a aplicar estos métodos, actualizando estas distribuciones iniciales en nuevas aplicaciones, con los resultados del paso anterior.

También existe la posibilidad de cambiar los modelos e interpretaciones a lo largo del análisis, mientras que en inferencia clásica se supone que las hipótesis y modelos se establecen antes de recoger los datos y no se cambian. Esto es poco razonable, pues dejar a los datos "hablar por sí mismos" es una idea básica en la modelización matemática, donde asumimos que los modelos son útiles para describir los datos, pero no son exactamente iguales a los datos y por tanto sería posible cambiar de modelo a lo largo del análisis (Pruzek, 1997; McLean, 2001).

Además, la influencia de la distribución inicial, depende del tamaño muestral y los posibles sesgos iniciales se corrigen en sucesivos experimentos, ya que el peso se vuelca sobre las verosimilitudes y se aumenta progresivamente el tamaño de la muestra (Lindley, 1993). Se recomienda también repetir el análisis con diferentes distribuciones iniciales e informar de las diferencias obtenidas en la distribución final (Zhu y Lu, 2004) y normalizar los procedimientos, usando distribuciones conjugadas, de modo que tanto la distribución inicial como la final pertenezcan a la misma familia de funciones (Cabriá, 1994).

Por otro lado, tampoco los métodos frecuenciales escapan de la subjetividad: el nivel de significación se establece arbitrariamente, por lo que unos mismos datos podrían ser estadísticamente significativos o no dependiendo del nivel de significación elegido. (Skipper, Guenter y Nash, 1970). La significación estadística deja de tener sentido cuando el tamaño de la muestra es tan grande que cualquier diferencia detectada, por pequeña que sea, permitirá rechazar la hipótesis nula. La definición de las variables, escalas de medición, pruebas de significación utilizadas, son otros ejemplos de las elecciones subjetivas del investigador. Más aún, la subjetividad es inevitable en la parte más importante de la inferencia, que es la interpretación de los resultados (Ayçaguera y Benavides, 2003).

Lo cierto es que la subjetividad no implica arbitrariedad, es inevitable en las ciencias sociales debido a la aleatoriedad inherente en sus variables y tiene un papel importante en la investigación científica. La comunidad científica acepta los diferentes hallazgos, estableciendo criterios metodológicos o de plausibilidad (Matthews, 1998).

¿Cómo Responde la Inferencia Bayesiana a las Necesidades del Investigador?

Varios trabajos sugieren que la inferencia bayesiana proporciona una respuesta más ajustada que la clásica a las necesidades del investigador (Lindley, 1993; Lecoutre, 1999; 2006).

En primer lugar, el sentido de la probabilidad en la estadística bayesiana es el mismo del lenguaje ordinario: medida condicional de incertidumbre asociada a la ocurrencia de un suceso,

cuando se asumen ciertos supuestos (Bernardo, 2003). Esta es la interpretación intuitiva- aunque incorrecta—que muchos científicos dan a las probabilidades frecuenciales asociadas al test de hipótesis, cuyos resultados interpretan inconscientemente en términos bayesianos (Falk, 1986; Gigerenzer, 1993; Rouanet, 1998; Lecoutre, 2006; Lecoutre, Lecoutre y Poitevineau, 2001; Haller y Krauss, 2002).

En consecuencia, la interpretación bayesiana de la inferencia parece ser más sencilla y natural que la frecuencial (Pruzek, 1997), además de proporcionar una base de toma de decisiones coherentes en situaciones de incertidumbre (Western, 1999). Además, la inferencia bayesiana proporciona un método totalmente general, debido a que su aplicación no requiere un tipo particular de distribución y no precisa deducir las distribuciones en el muestreo (Bernardo, 2003). A continuación analizamos la respuesta bayesiana a varias cuestiones que interesan al investigador.

Intensidad del efecto

Una recomendación para complementar los tests de hipótesis es estudiar la magnitud del efecto, pero una estimación puntual de los mismos es insuficiente, puesto que no tiene en cuenta el error muestral (Poitevineau, 1998). Un estudio de potencia sería recomendable para evitar conclusiones erróneas al concluir la ausencia de efecto cuando el resultado es no significativo (Cohen, 1990), pero el cálculo de la potencia no depende del valor observado del estadístico en la muestra y por tanto no es pertinente para interpretar un resultado particular, una vez recogidos los datos (Falk y Greenbaum, 1995). Los intervalos de confianza tienen la misma interpretación frecuencial que los contrastes, ya que sólo nos indican la proporción de intervalos calculados de la misma población con tamaño de muestra dado que cubrirían el valor del parámetro, pero no si el intervalo calculado lo cubre o no (Cumming, Williams, y Fidler, 2004).

El estudio del efecto y su magnitud aparecen en forma natural en los métodos bayesianos, que consideran el parámetro como variable aleatoria. La probabilidad de que dicho parámetro tome un cierto valor puede calcularse en la distribución final, pudiendo, por ejemplo utilizar frases como “la probabilidad de que el efecto supere tal magnitud es igual a 0,25”. El intervalo de credibilidad proporciona también los límites en los que el parámetro está incluido con una cierta probabilidad (Poitevineau, 1998; Lecoutre, 2006).

Contraste de hipótesis

El p-valor proporciona una probabilidad que no tiene utilidad para el investigador: la probabilidad de que obtuviésemos unos datos más extremos que los obtenidos, si repitiésemos innumerablemente el experimento y la hipótesis fuese cierta (Matthews 1998). Pero ningún investigador está interesado en repetir indefinidamente el mismo experimento y el fin de la investigación científica no es adoptar una decisión acerca de la veracidad de la hipótesis sino ajustar nuestro grado de creencia en la hipótesis que está siendo contrastada (Rozeboom, 1970).

Interpretar que, cuándo rechazamos la hipótesis nula, obtenemos un apoyo a la hipótesis de investigación (alternativa) no es adecuado, ya que un resultado significativo no indica la magnitud del efecto, por lo que la hipótesis estadística no informa sobre la significación práctica de los datos (Hager, 2000; Finch, Cumming, y Thomason, 2001). Esto puede producir situaciones en que rechazar una hipótesis nula no proporcione información nueva, ya que lo único que podemos deducir al rechazar una hipótesis es que hay un efecto, pero no en qué dirección o de qué magnitud (Falk y Greenbaum, 1995; Lecoutre, 1999).

Por el contrario, en la inferencia bayesiana podemos calcular probabilidades finales de la hipótesis y probabilidades de que el efecto tenga un cierto tamaño (Lindley, 1993). El método bayesiano es además comparativo. Compara la probabilidad del suceso observado bajo la hipótesis nula y bajo diferentes hipótesis alternativas (Lindley, 1993). Por otro lado, en algunas situaciones, como los ensayos de bioequivalencia, el interés se centra en verificar la hipótesis nula, es decir, esperamos es que los tratamientos sean equivalentes (Molinero, 2002). En estos casos el enfoque bayesiano resulta mucho más natural que el frecuentista, ya que tratamos de aceptar (no de rechazar) la hipótesis nula.

Probabilidades predictivas y replicación

Interpretar la significación estadística como seguridad en la replicabilidad de los datos no tiene base estadística (Falk, 1986; Gingerenzer, 1994; Cohen, 1994; Falk y Greenbaum, 1995; Pascual, García y Frías, 2000). La significación estadística no puede tomarse como evidencia de que la hipótesis de investigación sea cierta; ni proporciona la probabilidad de la hipótesis, por lo que no hay base para estudiar la replicación y tampoco nos proporciona evidencias verificables de replicación (Sohn, 1998).

En el método bayesiano podemos calcular la probabilidad de un suceso futuro, mediante la distribución predictiva dada en el denominador de la fórmula de Bayes, es decir, la media ponderada de la función de verosimilitud, ponderada por las probabilidades iniciales (Berry, 1995). Esto permite estudiar la posibilidad de replicación de nuestros resultados o estimar el tamaño necesario de muestra para que un futuro estudio sea concluyente (Lecoutre, 1996). Por supuesto, siempre que se cumplan los requisitos de precisión de los datos, adecuación de procedimientos y competencia de los investigadores (Sohn, 1998). Entendida correctamente, la replicabilidad se relaciona con la fiabilidad y consistencia de los datos, y la única forma de comprobarla es mediante sucesivos contrastes empíricos (Pascual, García y Frías, 2000).

Uso de la información previa

Mientras que los métodos frecuentistas consideran cada muestra como completamente nueva y no incorporan la información de estudios previos al actual, en el marco bayesiano es concebible una secuencia de experimentos articulados donde la información de cada uno se va incorporando al siguiente (Pruzek, 1997) e incluso reconoce la posibilidad de diferentes opiniones o conocimientos (Lindley, 1993).

Aunque es posible usar la inferencia bayesiana cuando no se tiene información previa sobre el parámetro, la característica más interesante es usar probabilidades iniciales “informativas” cuando sea posible, incluso para investigar el efecto de diferentes distribuciones iniciales supuestas. La idea central del enfoque bayesiano radica en actualizar los conocimientos probabilísticos que tenemos de un fenómeno, en función de la información disponible, por lo que resulta de gran interés en el meta-análisis, que combina información sobre diferentes estudios (Molinero, 2002).

Viabilidad Computacional de la Metodología Bayesiana

Un requisito para introducir los métodos de análisis de datos es la disponibilidad de programas de cálculo que faciliten su aplicación. En los últimos años son varios los investigadores que están desarrollando diversos programas, por lo que este enfoque se va introduciendo gradualmente en las Ciencias Sociales. Por ejemplo, Albert (1996) publicó una

serie de subrutinas de Minitab para el análisis bayesiano elemental, que se pueden descargar desde la página de Internet del autor (<http://bayes.bgsu.edu/>).

First Bayes (<http://www.tonyohagan.co.uk/1b/>) es un paquete preparado en la Universidad de Sheffield para la enseñanza de conceptos elementales Bayesianos, compatible con Windows, que se distribuye gratuitamente. Admite diferentes familias de distribuciones y calcula probabilidades iniciales finales y predictivas resultantes del análisis de modelos uniparamétricos, análisis de varianza de una vía y regresión (Lawrence, 2003).

Asimismo, el programa PAC (Lecoutre, 1996) permite analizar datos de diseños experimentales generales, incorporando métodos de comparación de medias y análisis de varianza univariante o multivariante, incluyendo medidas repetidas y covariables. El programa incorpora análisis frecuencial y bayesiano, con distribuciones a priori no informativas y conjugadas. Ha sido desarrollado por un equipo de investigación que trata de incorporar el análisis bayesiano en los métodos estadísticos más frecuentemente empleados en psicología. Una versión reducida se distribuye libremente desde la página web del equipo (<http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris.html>).

Para análisis más complejos Buggs (Bayesian inference Using Gibbs Sampling) es un software flexible interactivo y compatible con Windows, que permite cálculos complejos bayesianos, basados en la simulación (ver información en <http://www.mrc-bsu.cam.ac.uk/bugs/>). Ofrece facilidades en línea, tales como tutoriales, grupos de usuarios y ejemplos. El programa BACC (Bayesian Analysis Computation and Communication), desarrollado a partir de un proyecto financiado por la National Science Foundation en Estados Unidos, se va actualizando y ofrece recursos para el cálculo bayesiano, disponibles libres de costo. El énfasis se pone en la combinación de modelos y el desarrollo de distribuciones predictivas. Hay versiones disponibles para Matlab, S-PLUS y R, para sistemas Windows, Unix y Linux (<http://www2.cirano.qc.ca/~bacc/>).

Otros programas de cálculo bayesiano, algunos específicos y otros generales se van desarrollando y poniendo en Internet. Una lista de estos recursos puede verse en: http://www.mas.ncl.ac.uk/~ndjw1/bookmarks/Stats/Software-Statistical_computing/Bayesian_software/index.html/.

Viabilidad Didáctica de los Métodos Bayesianos Elementales

La introducción de una nueva metodología en psicología y educación requerirá de su difusión y comprensión por los posibles usuarios, es decir, del grado hasta el cuál seamos capaces de transmitir sus principales ideas en los cursos de estadística aplicada. En esta sección discutimos la posibilidad de enseñanza de este tema a alumnos de psicología o educación. Por ejemplo Iglesias y cols. (2000) sugieren la necesidad de que en los cursos de pregrado se introduzcan algunas ideas de inferencia bayesiana, junto con los de inferencia clásica, enfoque que ya se presentaba en algunos libros clásicos, como el de De Groot (1988) e indican que los temas de inferencia en estos cursos debieran ser, al menos, los siguientes:

Conceptos básicos fundamentales: Población, parámetro, muestra estadístico, función de verosimilitud, distribución inicial y final.

- Estimación puntual: Métodos clásicos y bayesianos
- Estimación por intervalos: Intervalos de confianza y de credibilidad.

- □ Contrastes de hipótesis: Tests clásicos y bayesianos, problemas de decisión múltiple.

En este sentido encontramos un número creciente de textos cuya comprensión no requiere grandes conocimientos matemáticos y en los que los elementos básicos de inferencia bayesiana se presentan contextualizados en ejemplos interesantes y familiares para los alumnos (por ejemplo Berry, 1995 o Albert y Rossman, 2001). Estos materiales pueden complementarse con las muchas referencias que explican en forma sencilla los rudimentos de inferencia bayesiana (ej. Ayçaguera y Benavides, 2003; Ayçaguera y Suárez, 1995).

Por otro lado, podemos encontrar en Internet también recursos didácticos que facilitan el aprendizaje de estos conceptos, tales como applets que permiten visualizar el teorema de Bayes o las distribuciones de probabilidad, o bien realizar los cálculos de distribuciones finales, para la inferencia de medias y proporciones con distribución a priori discreta o continua (ver, por ejemplo, la página de Jim Albert, <http://bayes.bgsu.edu/>). La mayoría de los autores citados en este apartado han incorporado los métodos bayesianos a su enseñanza y han indicado en sus escritos que los alumnos parecen comprender mejor la inferencia bayesiana que clásica. Encontramos también descripciones de experimentos concretos de enseñanza y sugerencias sobre la forma de llevarlos a cabo (Bolstad, 2002). Somos conscientes, sin embargo, de que esta posición es aún controvertida, incluso por educadores estadísticos prestigiosos (e.g. Moore, 1997). Esto es debido a que la investigación empírica sobre el aprendizaje real de los estudiantes en contextos de enseñanza naturales (dentro de los cursos de estadística) es aún muy incipiente.

Basándonos en todos estos antecedentes, nuestro grupo ha preparado un material elemental con la finalidad de ser utilizado en la enseñanza del teorema de Bayes e inferencia sobre media y proporción, en modalidad presencial o a distancia (Díaz, 2006). Consta de material escrito para el alumno, ejercicios resueltos, actividades y ejercicios de autoevaluación, así como subrutinas Excel de ayuda al cálculo. Siguiendo la sugerencia de Iglesias y cols. (2000), nos hemos concentrado más en transmitir conocimientos conceptuales que procedimentales (puestos que éstos vienen resueltos por el software) y enfatizando el lenguaje de comunicación con los alumnos, al contextualizar los conceptos en áreas de aplicación a psicología o de su experiencia cotidiana.

Con dicho material se realizó una primera experiencia de enseñanza con una duración de 15 horas a un total de 78 estudiantes de psicología, divididos en varios grupos. Nuestra primera impresión es positiva, al haberse conseguido la mayoría de objetivos de aprendizaje en una proporción alta de participantes. No obstante, continuamos revisando el material y pensamos realizar nuevas experiencias que mejoren nuestro conocimiento sobre la enseñanza de inferencia bayesiana.

Agradecimiento: Trabajo apoyado por el proyecto SEJ2004-00789, MEC, Madrid y FEDER y Grupo PAI FQM-126,. Junta de Andalucía

Referencias

- Albert, J. (1996). Bayesian computation using Minitab. Belmont, CA: Wadsworth.
- Albert, J.H. y Rossman, A. (2001). Workshop statistics. Discovery with data: A bayesian approach. Bowling Green, OH: Key College Publishing.

- Ares, V.M. (1999). La prueba de significación de la «hipótesis cero» en las investigaciones por encuesta. *Metodología de Encuestas*, 1, 47-68.
- Ayçaguera, L. y Suárez, P. (1995) ¿Qué es la inferencia bayesiana? *JANO* 1132, 1542. [On line]. Disponible en: http://www.atheneum.doyma.es/Socios/sala_1/lec06est.htm.
- Ayçaguera, L y Benavides, A. (2003). Apuntes sobre subjetividad y estadística en la investigación en salud. *Revista Cubana de Salud Pública*, 29(2), 170-173. On line disponible en: http://scielo.sld.cu/scielo.php?pid=S0864-34662003000200012&script=sci_arttext&tlng=es.
- Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, 2(1-2), 75 – 98.
- Batanero, C. y Díaz, C. (2006). Methodological and Didactical Controversies around Statistical Inference: Actes du 36ièmes Journées de la Societé Française de Statistique. CD ROM. Paris: Societé Française de Statistique.
- Batanero, C. y Díaz, C. (En prensa). Meaning and understanding of mathematics. The case of probability. En J.P Van Bendegen y K. Fraçois (Eds), *Philosophical Dimmensions in Mathematics Education*. New York: Springer .
- Bernardo, J.M. (2003). Bayesian Statistics. En R. Viertl (Ed.), *Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics*. Oxford, UK: UNESCO, On line: <http://www.uv.es/~bernardo/BayesStat.pdf>.
- Bernardo, J.M. A Bayesian Mathematical Statistics Primer. En Rossman, A., Chance, B. (eds.): *Proceedings of the Seventh International Conference on Teaching Statistics*. CD ROM. International Association for Statistical Education, Salvador (Bahia), Brazil (2006).
- Berry, D.A. (1995). *Basic statistics: A Bayesian perspective*. Belmont, CA: Wadsworth.
- Bolstad, W. M. (2002). Teaching bayesian statistics to undergraduates: Who, what, where, when, why, and how. En B. Phillips (Ed.). *Proceedings of the Sixth International Conference on Teaching of Statistics*. Ciudad del Cabo: IASE. CD ROM.
- Bolstad, W. (2004). *Introduction for Bayesian statistics*. New York: Wiley.
- Borges, A., San Luis, C., Sánchez, J.A. y Cañadas, I. (2001). El juicio contra la hipótesis nula: muchos testigos y una sentencia virtuosa. *Psicothema*, 13 (1), 174-178.
- Box, G.P. y Tiao, G.C. (1992). *Bayesian inference in statistical analysis*. Nueva York: Wiley.
- Cabriá, S. (1994). *Filosofía de la estadística*. Valencia: Servicio de Publicaciones de la Universidad.
- Cohen, J. (1990). Things I have learnt so far. *American Psychologist*, 45, 1304 – 1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49 (12), 997-1003.

- Cumming, G.; Williams, J. y Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics*, 3, 299-311.
- De Groot, M. H. (1988). *Probabilidad y estadística*. Delaware: Addison Wesley.
- Díaz, C. (2005). *Apuntes sobre inferencia bayesiana: Material didáctico*. Granada: La autora.
- Díaz, C. y de la Fuente, I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*, Volumen especial 2004, 161-167.
- Edwards, W., Lindman, H. y Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.
- 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 of a probabilistic misconception. *Theory and Psychology*, 5 (1), 75 – 98.
- Fidler, F. (2002). The fifth edition of the APA publication manual: Why its statistics recommendations are so controversial. *Educational And Psychological Measurement*, 62 (5), 749-770.
- Finch, S., Cumming, G., y Thomason, N. (2001). Reporting of statistical inference in the *Journal of Applied Psychology*: Little evidence of reform. *Educational and Psychological Measurement*, 61, 181-210.
- Frías, M.D., Pascual, J. y García, J.F. (2000). Tamaño del efecto del tratamiento y significación estadística. *Psicothema*, 12, supl. 2, 236-240.
- Gingerenzer, 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.
- Hacking, I. (1975). *The emergence of probability*. Cambridge, MA: Cambridge University Press.
- Hager, W. (2000). About some misconceptions and the discontent with statistical tests in psychology. *Methods on Psychological Research*, 5(1). On line. Disponible en <http://www.mpr-online.de>.
- Haller, H. y Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research*, 7(1). On line: <http://www.mpronline.de/issue16/art1/haller.pdf>.
- Harlow, L.L., Mulaik, S.A. y Steiger, J.H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

- Iglesias, P., Leiter, J., Mendoza, M., Salinas, V. y Varela, H. (2000). Mesa redonda sobre enseñanza de la estadística bayesiana. *Revista de la Sociedad Chilena de Estadística*, 16-17, 105-120.
- Lawrence, J. (2003). A quick introduction to First Bayes. Montreal: Mc Gill University: On line: <http://www.medicine.mcgill.ca/epidemiology/Joseph/pdf/First.Bayes.pdf> .
- Lecoutre, B. (1996). *Traitement statistique des données expérimentales: Des pratiques traditionnelles aux pratiques bayésiennes*. Paris : CISIA.
- Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute (Tome 58, Book 2)* (pp. 205-208). Helsinki, Finland: International Statistical Institute.
- Lecoutre, B. (2006). Training students and researchers in Bayesian methods for experimental data analysis. *Journal of Data Science*, 4, 207-232.
- 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.
- Lee, P. M. (2004). *Bayesian statistics. An introduction*. York, UK : Arnold.
- Levin, J. R., y Robinson, D. H. (1999). Further reflections on hypothesis testing and editorial policy for primary research journals. *Educational Psychological Review*, 11, 143 – 155.
- Lindley, D. V. (1993). The analysis of experimental data: The appreciation of tea and wine. *Teaching Statistics*, 15 (1), 22-25.
- Matthews, R. A. (1998). Facts versus factions: the use and abuse of subjectivity in scientific research. En J. Morris (Ed.), *Rethinking risk and the precautionary principle* (pp. 247-282). Oxford : Butterworth, 2000.
- McLean, A. (2001). Statistics in the catwalk. The importance of models in training researchers in statistics. En C. Batanero (Ed), *Training Researchers in the Use of Statistics*. Granada, Spain: International Association for Statistics Education and International Statistical Institute.
- Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal*, 5(1), 4 – 18.
- Molinero, A. (2002). El método bayesiano en la investigación médica. Asociación española contra la hipertensión arterial. On line: <http://www.seh-lelha.org/bayes1.htm>.
- Moore, D.S. (1997). Bayes for beginners? Some pedagogical questions. En S. Panchapakesan (Ed.), *Advances in Statistical Decision Theory* (pp. 3-17). Birkhäuser.
- Morrison, D.E., y Henkel, R.E. (Eds.). (1970). *The significance tests controversy. A reader*. Chicago: Aldine.

- O'Hagan, A. y Forster, J. (2004). *Bayesian Inference: Kendall's Advanced Theory of Statistics*. London: Arnold.
- Pascual, J., García, J.F. y Frías, M.D. (2000). Significación estadística, importancia del efecto y replicabilidad de los datos *Psicothema*, 12, Suplemento 2, 408-412.
- Poitevineau, J. (1998). *Méthodologie de l'analyse des données expérimentales: étude de la pratique des tests statistiques chez les chercheurs en psychologies: approches normative, prescriptive et descriptive*. Tesis doctoral., Universidad de Rouen.
- Pruzek, R.M. (1997). An introduction to bayesian inference and its applications. En L. L. Harlow, S. A. Mulaik y J. H. Steiger (Eds.). *What if there were no significance tests?* (pp. 287-318). Mahwah, NJ: Lawrence Erlbaum.
- Rindskopf, D.M. (1997). Classical and bayesian approaches. En L. L. Harlow, S. A. Mulaik y J. H. Steiger (Eds.). *What if there were no significance tests?* (pp. 319-334). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropos.
- Rouanet, H. (1998). Statistical practice revisited. En H. Rouanet et al. (Eds.) *New ways in statistical methodology* (pp. 29-64). Berna: Peter Lang.
- Rozeboom, W.W. (1970). The fallacy of the null hypothesis significance test. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 216 – 230). Chicago: Aldine.
- Skipper, J. K., Guenter, A. L., y Nass, G. (1970). The sacredness of .05: A note concerning the uses of statistical levels of significance in social sciences. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 155 – 160). Chicago: Aldine.
- Sohn, D. (1998). Statistical significance and replicability: Why the former does not presage the latter, *Theory & Psychology*, 8 (3), 291-311.
- Valera, A., Sánchez, J., Marín, F. y Velandrino, A.P. (1998). Potencia Estadística de la Revista de Psicología General y Aplicada (1990-1992). *Revista de Psicología General y Aplicada*, 51 (2)
- Valera, S., Sánchez, J. y Marín, F. (2000). Contraste de hipótesis e investigación psicológica española: Análisis y propuestas. *Psicothema*, 12(2), 549-582.
- Western, B. (1999), Bayesian analysis for sociologists: An introduction. *Sociological Methods & Research*, 28 (1), 7-34.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.

Zhu, M. y Lu, A.Y. (2004) The counter-intuitive non-informative prior for the Bernoulli family. *Journal of Statistics Education*, 12 (2), On line: <http://www.amstat.org/publications/jse/v12n2/zhu.pdf>

LAS AUTORAS

Carmen Días

Licenciada en Psicología, Diploma de Estudios Avanzados, Master Líneas de investigación: Inferencia bayesiana, metodología de investigación, didáctica de la estadística Departamento de Metodología de las Ciencias del Comportamiento Facultad de Psicología, Campus de Cartuja, 18071 Granada, España mc Diaz@ugr.es

Carmen Batanero

Licenciada en Matemáticas, Doctorado en Estadística Líneas de investigación: Didáctica de la estadística Departamento de Didáctica de las Matemáticas Facultad de Educación, Campus de Cartuja, 18071 Granada, España batanero@ugr.es

Datos de la Edición Original Impresa

Días, C y Batanero, C (2006, Diciembre). Validez y confiabilidad en la metodología cualitativa. *Paradigma*, Vol. XXVII, Nº 1, Diciembre de 2006. / 35-53