



Salud Pública de México

ISSN: 0036-3634

spm@insp.mx

Instituto Nacional de Salud Pública
México

Scotto, Manuel G; Tobías, Aurelio
Interpretando correctamente en salud pública estimaciones puntuales, intervalos de confianza y
contrastes de hipótesis
Salud Pública de México, vol. 45, núm. 6, noviembre-diciembre, 2003, pp. 506-511
Instituto Nacional de Salud Pública
Cuernavaca, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=10645611>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Interpretando correctamente en salud pública estimaciones puntuales, intervalos de confianza y contrastes de hipótesis

Manuel G Scotto, MSc, PhD,⁽¹⁾ Aurelio Tobías-Garcés, MSc, C Stat.⁽²⁾

Scotto MG, Tobías A.
Interpretando correctamente en salud pública estimaciones puntuales, intervalos de confianza y contrastes de hipótesis. *Salud Publica Mex* 2003;45:506-511.

El texto completo en inglés de este artículo está disponible en: <http://www.insp.mx/salud/index.html>

Resumen

El presente ensayo trata de aclarar algunos conceptos utilizados habitualmente en el campo de investigación de la salud pública, que en numerosas situaciones son interpretados de manera incorrecta. Entre ellos encontramos la estimación puntual, los intervalos de confianza, y los contrastes de hipótesis. Estableciendo un paralelismo entre estos tres conceptos, podemos observar cuáles son sus diferencias más importantes a la hora de ser interpretados, tanto desde el punto de vista del enfoque clásico como desde la óptica bayesiana. El texto completo en inglés de este artículo está disponible en: <http://www.insp.mx/salud/index.html>

Palabras clave: estimación puntual; prueba de hipótesis; intervalo de confianza; inferencia clásica; inferencia bayesiana

Scotto MG, Tobías A.

Accurate interpretation of point estimates, confidence intervals, and hypothesis tests in public health. *Salud Publica Mex* 2003;45:506-511.

The English version of this paper is available at: <http://www.insp.mx/salud/index.html>

Abstract

This essay reviews some statistical concepts frequently used in public health research that are commonly misinterpreted. These include point estimates, confidence intervals, and hypothesis tests. By comparing them using the classical and the Bayesian perspectives, their interpretation becomes clearer. The English version of this paper is available at: <http://www.insp.mx/salud/index.html>

Key words: point estimate; hypothesis test; on the up and up interval; classic inference; Bayesian inference

En los últimos años, en el campo de la salud pública, se ha reabierto¹⁻⁵ con creciente interés la controversia clásica⁶ sobre la combinación o comparación de las metodologías frecuentista y bayesiana en el diseño y análisis de datos clínicos y epidemiológicos. La incorporación de los métodos bayesianos ha abierto nuevas

perspectivas en este tipo de estudios, pero también ha provocado que muchos resultados sean erróneamente interpretados, como los intervalos de confianza y los valores de p . Esto ha provocado la aparición de gran número de trabajos^{4,7-12} con el objetivo de explicar la correcta utilización e interpretación de los resultados,

(1) Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal.

(2) Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe, España.

Fecha de recibido: 18 de septiembre de 2002 • Fecha de aprobado: 15 de abril de 2003

Solicitud de sobretiros: Aurelio Tobías Garcés. Departamento de Estadística y Econometría.

Universidad Carlos III de Madrid C/ Madrid 126, E-28903 Getafe (España).

Correo electrónico: atobias@est-econ.uc3m.es

dependiendo de cuál metodología, clásica o bayesiana, sea utilizada.

El objetivo de este ensayo es presentar de una manera formal los principios básicos de las metodologías frecuentista o clásica, y la bayesiana, revisando los conceptos teóricos de la estimación puntual, intervalos de confianza, y contrastes de hipótesis. Presentamos las diferencias fundamentales entre ambas metodologías, mostrando sencillos ejemplos que ayudan a clarificar las cuestiones más teóricas.

Principios de la inferencia clásica

En los problemas de inferencia estadística se trabaja casi siempre en el marco de un modelo probabilístico. En particular, en la inferencia clásica este modelo sirve para describir el proceso experimental o biológico a partir del cual se han generado los datos. El modelo clásico comprende un espacio de resultados, denominado espacio muestral, cuyos elementos resultan del conjunto de todos los posibles valores que puede tomar una variable aleatoria X . Las posibles funciones de probabilidad para X las designaremos por $f_{\theta}(x)$, siendo θ el vector de parámetros que clasifica dicha función. Entonces podemos decir que $\mathfrak{F} = \{f_{\theta}(x) : \theta \in \Theta\}$ forma una familia de distribuciones de probabilidad y constituye la parte central del modelo clásico. La inferencia clásica intenta responder a las siguientes preguntas: a) ¿son los datos recogidos compatibles con algún miembro de la familia \mathfrak{F} ?, y b) admitiendo que el modelo es válido ¿qué inferencias podemos hacer sobre el verdadero valor del vector de parámetros θ ? Es en este último punto donde se engloba la estimación paramétrica, puntual, o por intervalo.

La caracterización del modelo estadístico clásico refleja diferentes aspectos. El primero es que la información relevante que se utiliza para realizar inferencia es la obtenida por muestreo; y segundo, dichos procedimientos deben ser analizados a través de su comportamiento en un gran número de repeticiones, hipotéticas, efectuadas bajo las mismas condiciones. Este procedimiento se conoce como ‘principio de muestreo repetido’. Una de las facetas de este principio es que cualquier inferencia ha de basarse en el concepto ‘frecuentista’ de la probabilidad. Este concepto se fundamenta en la observación de la frecuencia relativa en un conjunto numeroso de experiencias independientes llevadas a cabo en las mismas condiciones. Otra de sus características importantes reside en la interpretación de los resultados con relación a la frecuencia con la que ocurren, es por ello que la distribución muestral de lo(s) estadístico(s) que define(n) el comportamiento a largo plazo del procedimiento inferencial juega un papel fundamental.

Estimación puntual

Supongamos que se quiere obtener una estimación puntual del parámetro θ . Llamaremos $\hat{\theta}(x)$ al estimador utilizado. Se puede considerar en términos generales que $\hat{\theta}$ será un buen estimador cuando tenga una probabilidad elevada de conducir a estimaciones ‘muy cercanas’ de θ . Por ejemplo, cuando para algún $\epsilon > 0$ pequeño $P(|\hat{\theta}(x) - \theta| < \epsilon) = 0.99$. Es decir, cuando a partir de un conjunto de observaciones se obtiene una estimación, $\hat{\theta}(x)$, lo único que se puede confirmar es una de estas dos aseveraciones:

1. Que $|\hat{\theta}(x) - \theta| < \epsilon$
2. Que ha ocurrido un acontecimiento improbable (por ejemplo, en este caso con $p = 0.01$).

Como el parámetro θ es desconocido, no sabemos en cuál de las dos situaciones nos encontramos. Desde el punto de vista de la inferencia clásica, recurrimos al principio de muestreo repetido, y afirmamos que si repitiésemos el cálculo de la estimación un gran número de veces la proporción de dichas estimaciones que verifican la primera condición estará cerca de 0.99. En otras palabras, el estimador $\hat{\theta}(x)$ es bueno porque en 99% de los casos conduce a estimaciones que difieren poco del verdadero valor del parámetro. Consideraciones análogas pueden hacerse en relación, por ejemplo, con los intervalos de confianza y con los contrastes de hipótesis.

Intervalos de confianza

Un intervalo de confianza se define mediante dos valores entre los cuales se encuentra el valor del parámetro con un cierto grado de confianza. El grado de confianza se refiere al procedimiento de construcción del intervalo, que se calcula con los datos muestrales y sus valores cambian de una muestra a otra. El grado de confianza no es la probabilidad de que un intervalo concreto incluya o no el valor del parámetro, ya que al ser el parámetro un valor fijo estará, o no, dentro de un intervalo concreto. El grado de confianza se refiere a la probabilidad de que, al aplicar repetidamente el procedimiento, el intervalo contenga el parámetro, es decir, expresa la proporción de intervalos que efectivamente incluyen el parámetro.

Por ejemplo, Burton y colaboradores⁷ proporcionan un ejemplo de la incorrecta interpretación que suele darse a los intervalos de confianza. Si el intervalo de confianza de 95% para una razón de mormos es [1.6-6.4], muchos investigadores interpretarían este resultado diciendo que hay una probabilidad de

95% de que el verdadero valor de la razón de momios está contenido en ese intervalo, lo que es incorrecto considerando la definición de intervalo de confianza descrita previamente.

Contrastes de hipótesis

Por lo que se refiere a los contrastes de hipótesis, cuando queremos contrastar una hipótesis nula simple H_0 contra una alternativa H_1 también simple, diciendo que una región crítica provoca un error de tipo I con probabilidad 0.05 y un error de tipo II con probabilidad 0.1, utilizando la interpretación frecuentista de la probabilidad, lo que se pretende decir es que si repitiésemos el contraste más veces, éste nos conduciría a rechazar incorrectamente H_0 en 5% de los casos y a aceptar de forma incorrecta H_1 en 10% de los casos.

Todas las probabilidades mencionadas hasta el momento (0.99, 0.1, y 0.05) hacen referencia a la *probabilidad previa* del estimador considerado. La probabilidad previa también recibe el nombre de pre-experimental, ya que se refiere a una situación anterior a la observación de los datos. La probabilidad previa es un importante criterio de evaluación de los procedimientos clásicos y no plantea problemas cuando se utilizan procedimientos que se repiten un gran número de veces. Lo que ciertamente es discutible es el empleo de la probabilidad previa en aquellas situaciones que sólo ocurren una vez y no se repiten.

Para ilustrar esta última situación, consideremos el siguiente ejemplo.¹³ Supongamos que a partir de una muestra queremos inferir sobre el parámetro λ de una distribución de Poisson. Para ello, utilizaremos el valor medio de los datos muestrales, \bar{X} , como estimador, sabiendo que bajo ciertas condiciones sigue una distribución normal, de valor medio λ y varianza λ/n . Con base en esta distribución, podemos decir que $\bar{X} \in (\lambda \pm 1.96\sqrt{\lambda/n})$ con una probabilidad a doble cola de 95%. Cualquiera de estas afirmaciones se refiere a la probabilidad previa, ya que está en juego el conjunto de valores posibles de \bar{X} . Lo que ahora nos preguntamos es ¿qué podemos decir de \bar{x} como estimación de λ ? Por ejemplo, si en una muestra de tamaño 16 obtenemos que $\bar{x}=16.8$, ¿qué podemos decir sobre la proximidad de este valor en relación a λ ? La respuesta que un estadístico clásico dará es que puesto que \bar{x} es un valor típico de \bar{X} , éste hereda las propiedades de la distribución muestral de \bar{X} . Esto nos revela que la metodología clásica no consigue medir la precisión efectivamente realizada, conocida como *probabilidad posterior*, o sea la proximidad entre \bar{x} y λ una vez observados los datos. Ejemplificando una vez más, si $\bar{x}=16.8$ es una media de una muestra concreta

de tamaño 16 de una población $N(\mu, l)$ lo que podemos decir es que el intervalo $(\bar{X} \pm 0.49)$ contiene μ con probabilidad 0.95, es decir que en 95% de las muestras, μ está contenido en $(\bar{x} \pm 0.49)$. Esto no significa que (16.8 ± 0.49) contenga μ , ya que podríamos haber obtenido una de 5% de las muestras para las cuales μ no está contenido en dicho intervalo. De este modo, no se consigue tener una idea de la mayor o menor plausibilidad de μ estar contenido en (16.8 ± 0.49) , es decir, de la probabilidad posterior del proceso de estimación por intervalo.

El siguiente ejemplo¹⁴ muestra los problemas que pueden aparecer cuando se hace un uso indiscriminado de la probabilidad previa. Consideremos que la variable $X \sim N(\mu, \sigma^2)$ es observada con el objetivo de contrastar $H_0: \mu=0$ contra $H_1: \mu=10$. La observación de X se puede hacer con dos instrumentos de medida: el primero, I_1 es poco preciso ($\sigma=10$) y el segundo, I_2 , es más preciso ($\sigma=1$). El observador usa I_1 o I_2 con probabilidades conocidas p y $1-p$, respectivamente, sabiendo siempre cuál instrumento utiliza. Para realizar el contraste H_0 contra H_1 podemos utilizar, entre otros, las siguientes pruebas, ambas de tamaño α :

Prueba 1. Utilizando el instrumento I_1 rechazamos H_0 cuando $X > k_0, k_0 = 10\phi^{-1}(1-\alpha)$ y utilizando el instrumento I_2 rechazamos H_0 cuando $X > k_1, k_1 = \phi^{-1}(1-\alpha)$ donde $\phi(x)$ es la función de distribución de la normal estándar. Es fácil verificar que la prueba 1 tiene tamaño α .

Prueba 2. Con el instrumento I_1 rechazamos siempre H_0 y con el instrumento I_2 rechazamos H_0 cuando $X > k_2, k_2 = \phi^{-1}(\frac{1-\alpha}{1-p})$, $\alpha > p$. El tamaño de la prueba 2 también es α .

Se puede demostrar que para varios valores de α y de p , la prueba 2 es más potente que la prueba 1. Por lo tanto, la doctrina clásica diría que la prueba 2 es mejor de las dos, si se tiene en cuenta la probabilidad previa. Si el observador se propone realizar una larga serie de ensayos no tiene ningún motivo que lleve a rechazar la recomendación de la doctrina clásica. El problema es que si el observador puede realizar una sola vez la experiencia (utilizando la prueba 2) a uno le extraña la idea de ignorar el resultado de no considerar la observación cuando utilizamos el instrumento I_1 . De hecho, es natural que al observador no le interese utilizar la probabilidad previa sino la precisión que obtendrá cuando se utilice cualquiera de los dos instrumentos de medida.

El papel central de la probabilidad previa en los procedimientos clásicos es consecuencia directa del principio de muestreo repetido. Este principio implica la transferencia, para los procedimientos estadísticos,

mismo cuando se tenga la intención de aplicarlos sólo una vez, de las propiedades que los mismos revelan en un conjunto ilimitado de repeticiones. Este es uno de los puntos de discordia entre clásicos y bayesianos, ya que puede haber ambigüedad a la hora de considerar este conjunto hipotético de repeticiones ilimitadas, pues uno puede imaginarse diferentes formas de repetir las experiencias y, en tal caso, no está claro cuál de las formas es la más adecuada para realizar inferencias sobre el parámetro.

Principios de la inferencia bayesiana

La inferencia bayesiana se fundamenta en el Teorema de Bayes sobre el cálculo de probabilidades condicionadas. La divergencia fundamental entre el modelo clásico y el modelo bayesiano es que el parámetro $\theta \in \Theta$, que en el modelo clásico es un escalar o vector desconocido pero fijo, pasa en la metodología bayesiana a ser considerado aleatorio. Para los bayesianos, todo lo que es desconocido es incierto y toda incertidumbre tiene que ser cuantificada en términos probabilísticos. Al mismo tiempo, admiten que la información inicial o *a priori* (anterior a la experiencia pero demasiado importante para ser ignorada) puede traducirse formalmente en una distribución de probabilidad para θ , que se conoce como distribución *a priori* que designaremos como $p(\theta)$. Por lo tanto, $p(\theta)$ proporciona el grado de credibilidad que una persona atribuye a los posibles valores de θ . Este aspecto de la distribución *a priori* es el más controvertido del método bayesiano, ya que mucha gente rechaza que la información inicial, que puede incluir los prejuicios del investigador, se incluya en un proceso de inferencia estadística. No obstante, este problema se soluciona por los bayesianos utilizando distribuciones *a priori* que no aporten información relevante sobre el parámetro.

Una vez cuantificada esta información inicial acerca del parámetro, las observaciones obtenidas por muestreo servirán para actualizar la información ya existente sobre el parámetro. De este modo, combinando la información inicial y la información muestral, utilizando el Teorema de Bayes se atribuyen nuevos grados de credibilidad a los valores del parámetro θ , que dan lugar a la distribución *a posteriori* $p(\theta|x)$. De aquí se deduce que todos los procedimientos de inferencia bayesiana se basan en la distribución *a posteriori*.

Estimación puntual bayesiana

Veamos ahora cómo resuelven los bayesianos el problema de la estimación puntual. El estimador bayesia-

no que normalmente se utiliza para obtener una estimación puntual es la moda de la distribución *a posteriori*, ya que es el valor más probable que puede tomar θ . En algunas situaciones, las estimaciones de máxima verosimilitud del modelo clásico y bayesiano son prácticamente idénticas. Esto ocurre, por ejemplo, cuando se elige una distribución *a priori* poco informativa, lo que implicará que la distribución *a posteriori* estará basada en la información obtenida por muestreo y en este caso las estimaciones para los dos modelos coincidirán. A pesar de esta circunstancia, la interpretación de los resultados, dependiendo de la metodología utilizada, es completamente diferente.

Intervalos de credibilidad

Con relación a los intervalos de confianza clásicos, en el modelo bayesiano se utilizan los llamados intervalos de credibilidad. Estos intervalos se construyen de la siguiente manera. Observado x , se quiere encontrar dos valores $[a_1, a_2]$ de tal manera que $\theta \in [a_1, a_2]$ con una probabilidad elevada, o sea $P(a_1 < \theta < a_2 | x) \geq 1 - \alpha$, siendo $1 - \alpha$ el nivel de credibilidad. Esta definición de intervalos de credibilidad presenta un inconveniente: hay muchos valores de a_1 y a_2 que cumplen esta propiedad. Se plantea entonces el problema de cómo elegir estos dos valores. Una buena solución es considerar el intervalo $C = \{\theta : p(\theta|x) > \gamma\}$ donde γ es el mayor número real que satisface $p(\theta \in C | x) \geq 1 - \alpha$. Esta regla da origen a un intervalo de amplitud mínima, de modo que cualquier punto excluido del intervalo no tenga mayor credibilidad que cualquier punto incluido en él. Estos intervalos se conocen como intervalos HPD (highest posterior density).

Diferencias entre la inferencia clásica y la bayesiana

Como ya vimos, la diferencia fundamental entre el modelo clásico y el bayesiano es que en este último los parámetros son considerados aleatorios, por lo que pueden ser cuantificados en términos probabilísticos. Por otro lado, es importante resaltar que la inferencia bayesiana se basa en probabilidades asociadas con diferentes valores del parámetro θ que podrían haber dado lugar a la muestra x que se observó. Por el contrario, la inferencia clásica se basa en probabilidades asociadas con las diferentes muestras x que se podrían observar para algún valor fijo, pero desconocido, del parámetro θ . En relación con la obtención de estimaciones puntuales para los parámetros poblacionales, en el caso del modelo clásico, la estimación se interpreta como el valor de θ , que hace más plausible

haber obtenido la muestra observada, mientras en el modelo bayesiano, la estimación será el valor de θ que, puesto que se ha observado x , sea más verosímil o más creíble.

Veamos cuáles son las diferencias que existen entre los intervalos de confianza clásicos y los intervalos de credibilidad bayesianos. En relación con la elección del α , no hay ninguna diferencia, elegimos siempre un valor pequeño. En los restantes aspectos no hay ninguna coincidencia. Visto que la distribución *a posteriori* es la distribución del parámetro, en el caso bayesiano, podemos hablar de la probabilidad de que el parámetro esté contenido en el intervalo de credibilidad. Esto contrasta con los intervalos de confianza clásicos, los cuales sólo pueden ser interpretados en relación con su probabilidad de cobertura, o sea, la probabilidad de obtener una muestra de manera que la región de confianza resultante contenga θ . Como se dijo, los intervalos clásicos son una medida de la probabilidad previa, mientras que los intervalos de credibilidad bayesianos son una medida de la probabilidad posterior. Esto se traduce en que una vez que se observa x , los bayesianos saben que con probabilidad como mínimo de $1-\alpha$, el verdadero valor del parámetro está contenido en el intervalo de credibilidad. Por lo tanto, y de manera contraria a lo que cabría esperar, la interpretación de los intervalos de confianza como se usa comúnmente es propia del método bayesiano y no del clásico, independientemente de la metodología utilizada para su cálculo.

Veamos un ejemplo. Supongamos que construimos el intervalo de manera que para las dos metodologías se obtenga $\bar{x} \pm 1.96\sigma/\sqrt{n}$. Mientras que para un bayesiano $P(\bar{x} - 1.96\sigma/\sqrt{n} < \theta < \bar{x} + 1.96\sigma/\sqrt{n} | x)$ tiene una interpretación directa como probabilidad, representando por lo tanto la probabilidad posterior, para un clásico, antes de obtener una muestra dicho intervalo representa la probabilidad de cobertura de θ por un intervalo aleatorio (probabilidad previa). Después de recoger la información por muestreo lo único que un clásico podrá decir es que, o efectivamente el verdadero valor del parámetro está en ese intervalo, o bien, que ocurrió un suceso de probabilidad 0.05.

Finalmente, veamos cuáles son las diferencias fundamentales con relación a los contrastes de hipótesis. Supongamos que se quiere contrastar si $H_0: \theta \in \Theta_0$ respecto $H_1: \theta \in \Theta_1$. Como ya se explicó anteriormente, siguiendo la metodología clásica este contraste se llevaría a cabo delimitando una región de rechazo V en el espacio muestral, que sería evaluada a través de las probabilidades de los errores de tipo I (probabilidad de rechazar la hipótesis nula cuando es cierta) y de tipo II (probabilidad de no rechazar la hipótesis nula

cuando es falsa). Estas probabilidades reflejan la probabilidad previa. Desde el punto de vista bayesiano el problema del contraste de hipótesis es encarado de manera más directa, calculando las probabilidades *a posteriori*, $P(\Theta_0 | x)$ e $P(\Theta_1 | x)$ decidiendo después de compararlas. Por ejemplo, se puede optar por la hipótesis nula si $P(\Theta_0 | x) > P(\Theta_1 | x)$. Estas probabilidades reflejan la probabilidad posterior del procedimiento. Por lo tanto, en el contexto de los contrastes de hipótesis, se puede afirmar que el resultado de la inferencia bayesiana no es la aceptación o rechazo de las hipótesis, pero sí la alteración de la credibilidad que es atribuida a cada una de ellas.

Conclusiones

En el presente ensayo se ha ofrecido una visión global de los conceptos básicos para hacer inferencia, estimaciones puntuales, intervalos de confianza, y pruebas de hipótesis, bajo el prisma de la metodología clásica y de la metodología bayesiana. Aunque es importante recalcar que ni el método inferencial clásico ni el bayesiano prescinden de un abordaje general que incluya los criterios de causalidad, el método adecuado, la calidad de los datos, y el contexto del problema de salud bajo investigación.

Las inferencias generadas a partir de la metodología bayesiana son más informativas y fáciles de interpretar que las realizadas a partir de la metodología clásica, proporcionando respuestas más directas a muchas de las cuestiones planteadas por los investigadores en el campo de la epidemiología y de la salud pública.^{4,8-10} Pero encontramos que todavía es bastante habitual que, tanto epidemiólogos como clínicos interpreten los intervalos de confianza calculados por métodos clásicos utilizando el concepto bayesiano de la probabilidad subjetiva.⁷ Este hecho es debido principalmente a dos motivos: en primer lugar, la escasa repercusión que los métodos bayesianos han tenido, tanto en los libros de texto más populares en el ámbito de la bioestadística,¹⁵⁻¹⁸ como en la formación de los investigadores. Por otro lado, la total inviabilidad de los programas estadísticos de carácter general de mayor difusión en el ámbito de la investigación médica (como Stata, SAS, o SPSS) respecto de la aplicación de técnicas estadísticas bayesianas.

En este sentido, sería deseable un cambio mucho más progresivo en la manera de concebir y analizar los estudios epidemiológicos actuales, de tal manera que la metodología bayesiana se convierta en el modo natural de concebir los análisis estadísticos con fines inferenciales. Este cambio de pensamiento ha coincidido con el desarrollo de los métodos avanzados

de simulación (Monte Carlo –MC– o Markov Chain Monte Carlo –MCMC–)¹⁹ y su implantación en el *software* estadístico para análisis bayesiano, principalmente WinBugs,²⁰ haciendo posible que la aplicación de la metodología bayesiana sea cada vez más habitual en el campo de la salud pública.^{21,22} Pero mientras un progresivo cambio de mentalidad se va produciendo, coincidimos en que conceptos de la metodología clásica, tal como el valor de p , no deben ser descartados radicalmente.^{22,23} Por ello compartimos con Bacallao² la recomendación de que mientras el pensamiento bayesiano va desplazando al clásico, conviene utilizar las pruebas estadísticas habituales únicamente cuando se pueda garantizar un alto grado de plausibilidad de la hipótesis nula a contrastar.

Agradecimientos

Agradecemos a los dos revisores anónimos sus valiosos comentarios a versiones previas del manuscrito, así como a los editores por su gentileza a la hora de considerar este manuscrito en formato de ensayo para su publicación.

Referencias

1. Etzioni RD, Kadane JB. Bayesian statistical methods in public health and medicine. *Annu Rev Public Health* 1995;16:23-41.
2. Bacallao J. La perspectiva exploratorio-confirmatoria en las aplicaciones biomédicas de la estadística: dos diálogos (I). *Bayesianismo frente a frecuentismo: sus respectivas implicaciones prácticas en relación con el análisis de datos*. *Med Clin* 1996;107:467-471.
3. Bacallao J. La perspectiva exploratorio-confirmatoria en las aplicaciones biomédicas de la estadística: dos diálogos (II). *Consideraciones críticas acerca de las pruebas de significación*. *Med Clin* 1996;107:539-543.
4. Burton PR, Gurrin LC, Campbell MJ. Clinical significance not statistical significance: A simple Bayesian alternative to p values. *J Epidemiol Community Health* 1998;52:318-323.
5. Silva LC, Muñoz A. Debate sobre métodos frecuentistas vs métodos bayesianos. *Gac Sanit* 2000;14:482-494.
6. Rothman KJ. A show of confidence. *N Engl J Med* 1978;299:1362-1363.
7. Burton PR. Helping doctors to draw appropriate inferences from the analysis of medical studies. *Stat Med* 1994;13:1699-1713.
8. Goodman SN. Of p -values and Bayes: A modest proposal. *Epidemiol* 2001; 12: 295-297.
9. Silva LC, Benavides A. El enfoque bayesiano: otra manera de inferir. *Gac Sanit* 2001;15:341-346.
10. Bayarri JM, Cobo E. Una oportunidad para Bayes. *Med Clin* 2002;119:252-253.
11. Alamo F, Vázquez FJ, Rodríguez JC. Herramientas para la investigación biomédica: la perspectiva bayesiana (I). *Med Clin* 2002;119:265-268.
12. Alamo F, Vázquez FJ, Rodríguez JC. Herramientas para la investigación biomédica: la perspectiva bayesiana (II). *Med Clin* 2002;119:269-272.
13. Murteira JF. *Estatística: Inferência e decisão*. Lisboa: Imprensa Nacional, Casa da Moeda, 1988.
14. Cox DR. Some problems connected with statistical inference. *Ann Math Stat* 1958;29:357-372.
15. Colton T. *Statistics in medicine*. Boston: Little Brown, 1974.
16. Armitage P, Berry G. *Statistical methods in medical research*. Oxford: Blackwell, 1987.
17. Bland M. *An introduction to medical statistics*. Oxford: Oxford University Press, 1987.
18. Campbell MJ, Machin D. *Medical statistics: A commonsense approach*. Chichester: John Wiley, 1993.
19. Casella G, George EL. Explaining the Gibbs sampling. *Am Stat* 1992;46:167-174.
20. Spiegelhalter D, Thomas A, Best N, Gilks W. *BUGS Bayesian inference using Gibbs sampling version 0.50*. Cambridge: MRC Biostatistics Unit, 1995.
21. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. *Methods in health service research. An introduction to Bayesian methods in health technology assessment*. *Br Med J* 1999;319:508-512.
22. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. *Bayesian methods in health technology assessment: A review*. *Health Technol Assess* 2000; 4:1-130.
23. Cobo E, Campbell MJ. Interpreting results of observational research. *P values are still useful*. *Br Med J* 1994;309:1439.
24. Weinberg CR. It's time to rehabilitate the p -value. *Epidemiol* 2001;12:288-290.