

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

Jorge Bacallao

Instituto Superior de Ciencias Médicas de La Habana. Cuba.

Platón es quizás el primero y, sin duda, el más relevante dialoguista de la historia. La filosofía de su maestro Sócrates, verdadero compendio de las ideas de la época sobre el hombre y la naturaleza, nos ha sido transmitida en sus famosos *Diálogos*. Sus contemporáneos preferían divulgar sus ideas, siguiendo la tradición de Euclides, en el estilo teorema de deducción y prueba.

Posteriormente, el diálogo fue con frecuencia elegido como estilo de comunicación por filósofos (*Los Moralia*, de Plutarco; *Los Diálogos de los Muertos*, de Luciano de Samosata), políticos (*La República*, de Cicerón; *El arte de la guerra*, de Maquiavelo) y religiosos (*Los Diálogos sobre la Elocuencia*, del obispo Fenelon, y los *Diálogos*, del reformista Ulrich de Hutten).

En la ciencia, Kepler, con su famosa *Physica Coelestis*, y Galileo, con su célebre *Diálogo sobre los dos principales sistemas del mundo*, que colmara el furor de Scipione Chiaramonti y los filósofos de Pisa, y sellara su destino ante los tribunales de la Inquisición, han sido los dos más destacados exponentes de la forma dialogal.

En la época actual, el formato euclideano sigue siendo la forma de presentación habitual en las revistas matemáticas teóricas, mientras que en las revistas aplicadas, particularmente en el campo de la biomedicina, aquel estilo devino en la variante «métodos-resultados-discusión» que algunas publicaciones exigen como diseño formal obligado.

El diálogo, casi en desuso en la actualidad*, tiene la ventaja de prestarse a la presentación de puntos de vista dispares, y este hecho lo hace particularmente apropiado para la exposición de asuntos polémicos.

Uno de estos asuntos lo es sin duda el análisis de datos en las investigaciones biomédicas, en donde los esquemas clásicos basados en las pruebas de significación (Fisher) y en la verificación de hipótesis (Neumann-Pearson) resultan, si no inapropiados, por lo menos insuficientes para enfrentar una gran cantidad de problemas muy frecuentes, especialmente en la clínica y la epidemiología.

Lo que sigue a continuación son dos diálogos en los que se abordan, sin la pretensión de agotarlos, algunos de los problemas metodológicos más candentes del análisis estadístico aplicado a las ciencias médicas. Sus protagonistas representan, respectivamente, los puntos de vista de un usuario crítico de las técnicas, de un estadístico clásico que defiende la postura ortodoxa difundida por los manuales, los paquetes estadísticos y los cursos tradicionales a profesionales

*Notables excepciones son los *Diálogos sobre el Conocimiento* y el *Diálogo sobre el Método*^{1,2} de Paul Feyerabend.

Correspondencia: Dr. J. Bacallao.
Instituto Superior de Ciencias Médicas de La Habana.
Calle 146. N.º 3102, esq. 31. Cubanacán. Playa.
La Habana 11600.

Manuscrito aceptado el 10-3-1996

Med Clin (Barc) 1996; 107: 467-471

no matemáticos, y el de quien podríamos denominar un estadístico pragmático, que suscribe un mayor relativismo epistemológico, y a quien se ha escogido como principal vocero de la perspectiva que hemos llamado *exploratorio-confirmatoria*, ella misma, por supuesto, también polémica**.

Tras esta introducción, no creemos necesario anticipar a quién es más afín el punto de vista personal del autor.

I. Bayesianismo frente a frecuentismo: sus respectivas implicaciones prácticas en relación con el análisis de datos

SIMPLICIO: Todavía encuentra uno por ahí mucha gente que quiere publicar cualquier hallazgo sin averiguar si es realmente significativo o no. Recuerdo haber procedido así durante mucho tiempo, hasta que conseguí, primero entender, y luego aceptar el concepto de significación estadística.

NORMANCIO: Esa actitud a que te refieres, Simplicio, demuestra que los estadísticos aún tenemos que librar una batalla sin cuartel contra la costumbre de querer descansar en juicios subjetivos y no recurrir a la imparcialidad del dato.

S: Claro, que el dato no siempre es imparcial...

N: Por supuesto, pero demos por sentado que lo es y que ha sido obtenido de acuerdo a las buenas normas de la práctica científica, porque de lo contrario tendríamos que dudar de lo que constituye el cimiento del método experimental.

En verdad, Simplicio, las pruebas de significación han tenido el efecto, por un lado, de proteger a la ciencia de la arrogancia de los investigadores, y por otro, de proteger a los investigadores del riesgo del descrédito.

S: Me siento inclinado a creer que es así, pero me gustaría que concretaras el argumento.

N: Protege a la ciencia porque la libra de la arbitrariedad del juicio subjetivo. La significación estadística es un criterio universal para establecer la relevancia de un hallazgo o una observación, que podemos compartir o no, pero que introduce un elemento de entendimiento común. Protege al investigador porque pone de relieve que a nuestras creencias y a nuestras decisiones se asocia siempre un margen de incertidumbre, que es esencialmente inevitable.

**Estos personajes son los homólogos hispano-parlantes de otros dialoguistas como los que protagonizaron el debate reseñado en el artículo de Salsburg (1990)³.

S: Y que tiene su origen –corrígeme si me equivoco– en lo que constituye la base de toda la inferencia estadística, o sea, en no poder examinar todo el universo que nos interesa y vernos obligados a examinar sólo una parte de él.

N: Aprendiste bien.

LIPIDIO: Sin embargo, en lo que acabas de decir hay una inconsistencia que muy pocos se han detenido a considerar: salvo en los casos excepcionales en que tratamos con poblaciones finitas y reales, nunca muestreamos en un sentido estricto, a pesar de que en toda la estadística inferencial el muestreo probabilístico, y particularmente, el muestreo aleatorio, son condiciones permanentes.

S: ¿Y cuál es el origen de la inconsistencia?

L: El hecho de que en la investigación médica, tanto en la clínica como en la epidemiología, trabajamos con poblaciones virtuales, que no son muestreables. Los sujetos cuyos atributos define mi diseño experimental –por ejemplo, diabéticos insulino dependientes con menos de 10 años de evolución, tratados con un producto hipoglucemiante X a la dosis d – no forman parte de ninguna población con existencia real que pueda ser muestreada. Los resultados del acto inferencial que sigue al diseño se proyectan tanto hacia el pasado como hacia el futuro y obligan a recurrir gratuitamente al principio de que tanto uno como el otro deben asemejarse al presente. El recurso a este principio extrínseco por excelencia, que implica el paso de una población real a una virtual, plantea no sólo un problema de validez, sino que remite a los fundamentos de la inducción. Creo que todo el esquema inductivo que hay detrás de las pruebas de significación podría aceptarse sólo si consentimos en imprimirle un carácter autocorrectivo, en el sentido de que cualquier conclusión es provisoria, y cualquier prueba, un recurso modificador del grado de confirmación de las hipótesis. Pero en tal caso, hay formas mucho más naturales de proceder que las pruebas de significación.

S: Si no entendí mal, lo que quieres decir es que una prueba de significación no deja definitivamente resuelto el problema subyacente de verificación de hipótesis.

L: Así es, pero hay más aún: apenas tiene efecto sobre la necesidad de reiterarla. Ninguna de las convicciones que hoy nos atreveríamos a sostener acerca de la asociación entre un factor de exposición y un daño, o acerca de la superioridad de un tratamiento sobre otro, tiene su fundamento en una única prueba de significación, independientemente del valor de p asociado a ella. Muchos efectos reiterados y sistemáticos en la misma dirección, aunque no sean significativos, tienen en la práctica mayor capacidad persuasiva que uno solo muy significativo.

S: Confieso que me he extraviado en tu argumento, pero no creo que tu intención sea poner en duda el fundamento de las pruebas estadísticas.

L: Preferiría volver sobre el asunto en otra ocasión. Por el momento sólo querría decir que, ciertamente, el uso indiscriminado de los tests estadísticos y las pruebas de significación ha fomentado la expectativa de que es posible dar una respuesta positiva o negativa a preguntas tales como: ¿las variables X e Y están asociadas?, o ¿el tratamiento A es mejor que el B? Es muy fácil traducir estos

interrogantes a los términos de un problema de prueba de hipótesis, diseñar un experimento simple y practicar una prueba de significación. Para mí cualquiera de estos interrogantes es, en el fondo, equivalente a la pregunta de si un hombre es o no es alto, o es o no es veloz, que no pueden tener una única respuesta a menos que nos remitamos explícitamente a un criterio extrínseco. Si nos sirviésemos de analogías como ésta para considerar los problemas científicos en la clínica, la epidemiología, o incluso, las ciencias básicas médicas, creo que no hay modo de evitar ser escépticos en cuanto a la pertinencia de los tests y del concepto de significación estadística para el tratamiento de aquellos problemas.

N: En el fondo lo que quieres decir es que la respuesta a cualquiera de aquellos interrogantes plantea un dilema de interpretación. Justamente en ello radica la utilidad fundamental de las pruebas de significación: en que constituyen una salvaguarda contra la diversidad de interpretación y contra las teorizaciones audaces, o para decirlo al estilo de Fleiss^{4,5} contra el peligro de que lo que para mí es una diferencia sustantiva, para ti pueda ser una diferencia trivial.

L: Mucho más importante es que estimulemos el sentido crítico en la ciencia con el fin –y cito a Poole⁶– «de salvaguardarnos a nosotros y a otros de los peligros y de los ritos de la significación estadística».

S: Es desconcertante oír de una fuente autorizada el calificativo de rito para lo que yo había asimilado como una herramienta habitual de mis investigaciones. ¿Hay alguna opción que sustituya a la significación estadística?

N: La opción más importante es la estadística bayesiana, pero su éxito ha sido relativo, entre otras cosas, porque no ha logrado desplazar al enfoque clásico en el tratamiento de los problemas más comunes que plantea a diario la investigación biomédica.

S: Algo he oído, pero quisiera que me resumieras lo esencial del punto de vista bayesiano.

N: A mi modo de ver, hay tres elementos esenciales que lo caracterizan. El primero es su carácter decisional: el enfoque bayesiano, al igual que el enfoque de Neumann-Pearson, supone la existencia de dos hipótesis contendientes (la hipótesis nula y la alternativa) definidas sin ninguna ambigüedad, y de un criterio para la elección de una de estas hipótesis. Aceptar una de ellas como resultado del análisis estadístico no equivale a declararla verdadera, sino a optar por considerar que lo es; no implica creer en su verdad, sino decidir actuar como si fuera verdadera. Este estilo inferencial no representa para sus defensores –como tampoco para los defensores del enfoque de Neumann-Pearson– un razonamiento inductivo, sino un comportamiento inductivo.

El segundo elemento es que esta elección no es concebible si no se posee una creencia a priori acerca de la credibilidad relativa de cada una de las hipótesis. Dicho de otro modo, el enfoque bayesiano no nos permite formarnos una opinión únicamente a partir de indicios contenidos en los datos empíricos, sino que nos señala tan sólo cómo debe ser modificado nuestro juicio precedente.

El tercer elemento es que cada posible decisión errónea tiene un costo...

S: Algo así como los errores de primero y segundo tipos en el criterio de Neumann-Pearson.

N: Con la diferencia de que estos costos son introducidos de manera explícita en el algoritmo de decisión.

S: Creo captar la idea central, pero confieso que me resulta un poco abstracta tu exposición.

L: Hay una analogía formal con el diagnóstico clínico que te ayudará a comprender. Las condiciones de SANO o ENFERMO corresponden a las hipótesis nula y alternativa, respectivamente. La sensibilidad, que, como sabes, es la proporción de enfermos en que un test diagnóstico t resulta positivo, tiene su homólogo en la potencia de un test estadístico para detectar una hipótesis alternativa, y la especificidad (que es la proporción de sujetos sanos en que el test resulta negativo), se identifica con 1 menos la probabilidad del error de primer tipo.

Ahora bien, para el clínico no bastan la sensibilidad y la especificidad. Necesita, además, conocer los valores predictivos, es decir, la proporción de sujetos realmente enfermos entre todos los casos con test positivo (valor predictivo positivo: VPP) y la proporción de sujetos sanos entre todos los tests negativos (valor predictivo negativo: VPN). VPP y VPN pueden calcularse a través de la sensibilidad y la especificidad, sólo si se conoce además la prevalencia de la enfermedad, o su frecuencia relativa dentro de cierta población de interés.

El teorema de Bayes, piedra angular de todo el enfoque bayesiano, es esencialmente una expresión para el tránsito de la prevalencia (que es una especie de creencia inicial en la enfermedad del sujeto) a los valores predictivos (que son una creencia final, que ha sido modificada por el resultado del test, o sea, por la evidencia empírica). La influencia de esta creencia inicial es enorme. Te pondré sólo un ejemplo que puedes encontrar en Boen (1993)⁷. Otro similar y aún más clásico aparece en el libro de Parzen (1960)⁸.

Consideremos un test diagnóstico con sensibilidad y especificidad iguales a 0,99 (excepcionalmente altas, por supuesto) y supongamos que quisiéramos aplicarlo para el pesquijaje de una enfermedad en un país de 25 millones de habitantes, entre los cuales se calcula que hay aproximadamente 100.000 enfermos (prevalencia de 1 entre 250). El valor predictivo positivo calculado a través del teorema de Bayes sería de 0,284, o sea que, de cada mil resultados positivos, encontraríamos 716 falsos positivos, lo que representa una proporción intolerablemente alta, a pesar de las obvias bondades del test.

Ya te habrás dado cuenta de que todo ha sido efecto de la baja prevalencia (que los estadísticos designamos con la expresión «baja probabilidad a priori»), y de que, por supuesto, el resultado sería bien diferente si el test se aplicase a una población seleccionada en que esta probabilidad a priori, que expresa nuestra creencia previa en la condición patológica de los sujetos, fuese mayor. Por ejemplo, si pudiéramos suponer que el 50% de los sujetos examinados está enfermo, bajo las mismas condiciones anteriores, este test arrojaría valores de VPP y VPN de 0,99.

S: No alcanzo a ver qué interpretación podrían tener los valores predictivos (VP) dentro del lenguaje o la nomenclatura bayesiana. Tampoco a qué concepto corresponden estos VP en el contexto de las pruebas de hipótesis.

L: Los valores predictivos representan el nuevo nivel de nuestra creencia en la condición de un sujeto. Este nuevo nivel de creencia resulta del efecto que tiene el conocimiento del resultado del test sobre nuestra creencia inicial representada por la prevalencia.

S: Es decir que, ¿el resultado del test representa para nuestra decisión o nuestro diagnóstico clínico lo que los datos para nuestra preferencia por una u otra hipótesis?

L: En efecto. Ahora bien, repara en que el enfoque clásico de las pruebas de hipótesis no permite incorporar este conocimiento previo. Un resultado se acepta como significativo cuando tiene una probabilidad muy baja de ocurrir bajo la hipótesis nula, en donde se ha aceptado convencionalmente que una «probabilidad muy baja» es una probabilidad igual o menor que 0,05. Para nada se toma en cuenta la probabilidad que ese mismo resultado tendría para una alternativa fija dada. El enfoque de las pruebas de hipótesis de Neumann-Pearson si toma en cuenta esta probabilidad, pero también pasa por alto la creencia a priori en la hipótesis. Actuar de este modo equivale a diagnosticar a un sujeto cualquiera como enfermo luego de un resultado positivo con un test muy específico, independientemente de si existen o no antecedentes clínicos que permitan sospechar su enfermedad, lo cual significa derrochar o usar ineficientemente la información que se posee.

S: No he visto dónde y cómo entran los costos en todo este proceso.

N: Éste es uno de los puntos débiles del enfoque bayesiano. ¿Crees que tiene algún sentido estimar el costo de un error científico? Los bayesianos calculan el costo esperado asociado a cada una de las decisiones posibles, utilizando las probabilidades a posteriori (los VP del diagnóstico en la analogía de Lipidio) y adoptan la decisión que tenga el menor costo o la mayor utilidad posible. Así incorporan los costos al proceso.

S: ¿Y no podría eliminarse el problema atribuyendo costos iguales a las dos posibles decisiones erróneas? Si entiendo bien todo quedaría reducido entonces a optar por la hipótesis con una mayor probabilidad a posteriori, o para emplear el lenguaje que empleó Lipidio, la que fue beneficiada con nuestra mayor creencia final.

N: Desde luego, pero subsiste la mayor de todas las dificultades, que es la imposibilidad de encontrar un equivalente del concepto de prevalencia en el contexto de las pruebas de hipótesis. La probabilidad a priori de un diagnóstico puede identificarse con su frecuencia relativa en una población dada, pero no ocurre lo mismo en el caso de una hipótesis. Hablar de la probabilidad de una hipótesis no posee, en rigor, ningún sentido. Una hipótesis es verdadera o falsa y me parece absurdo recurrir a la noción de probabilidad como expresión de nuestra ignorancia o incertidumbre acerca de su condición.

L: Sin embargo, para los bayesianos, que suscriben el punto de vista subjetivo en la interpretación de la probabilidad, la noción de grado de creencia en la probabilidad de una hipótesis no sólo es aceptable, sino que constituye la idea central del cálculo de probabilidades. A mí me resulta atractiva la idea de que el papel de las observaciones empíricas consiste en modificar nuestras creen-

cias subjetivas, pero creo que por el momento se corre el gran riesgo de que las hipótesis planteadas por un investigador resulten confirmadas con gran facilidad a través de cualquier procedimiento que se apoye en una probabilidad inicial subjetiva, a la que, lógicamente, el investigador tendría tendencia a atribuir un alto valor.

S: Dices que «por el momento» existe ese riesgo. No veo cómo en el futuro pueda eludirse o dejar espontáneamente de existir.

L: Cuando el interés de los investigadores sea medir efectos y no aceptar y rechazar hipótesis, será mucho más fácil, y podrá realizarse sobre bases mucho más objetivas, la representación de la creencia previa a través de una función de probabilidad. Dicho de otro modo, los resultados de muchos ensayos clínicos diseñados, por ejemplo, para comparar dos medicamentos, dejarán de ser afirmaciones en el sentido de que existen o no diferencias significativas, para convertirse en aproximaciones cada vez más precisas a la medición de la diferencia entre los efectos de ambos medicamentos y no será difícil formular un modelo probabilístico para la distribución a priori de esa diferencia, que exprese, no una creencia individual, sino el consenso de una comunidad científica.

La forma estereotipada y rutinaria en que se aplica la investigación moderna, aun con poca o ninguna cultura en los fundamentos filosófico-metodológicos del bayesianismo, no tiende a alentar un cambio. Sin embargo, ya hay indicios (el metaanálisis^{9,10} es el más elocuente) de que el interés de los investigadores se va desplazando hacia la medición de efectos, y que hay una tendencia a abandonar el esquema estrecho de la significación estadística.

N: Me parece poco probable que este punto de vista llegue a imponerse. Efron (1986)¹¹ exponía una serie de razones para fundamentar su pronóstico de que al enfoque clásico le aguarda todavía una larga vida, y al bayesianismo una larga espera antes de convertirse en instrumento habitual de la investigación. Yo veo además otra razón, y es que la ciencia no se conforma con ser un inventario de mediciones, aunque éstas tengan cada vez una mejor aproximación, sino que aspira a conclusiones que se concretan en proposiciones científicas. Por eso, el conocimiento científico es mucho más afín al lenguaje de las pruebas de hipótesis. Como te decía al principio, el nivel de significación puede ser tan subjetivo o arbitrario como una probabilidad a priori, pero constituye un elemento de entendimiento común.

L: Ninguna conclusión o proposición del tipo «el tratamiento A es mejor que el B» puede ser más informativa o más científica que una buena medición de la diferencia entre los efectos de A y B, salvo por un hábito inveterado de economía intelectual al que no pienso que estemos fatalmente condenados.

S: Pero, ¿qué nos queda a corto plazo si rechazas el esquema clásico y objetas el modelo bayesiano como impracticable en un futuro inmediato?

L: Quedan algunas opciones que discutiremos oportunamente. Por lo pronto, yo quisiera citar a Gosset (1965)¹², a quien seguramente conoces por su seudónimo Student: «... el test de significación *no prueba* por sí mismo que la muestra no ha sido obtenida de la población por

simple azar, aun cuando el valor de *p* sea muy pequeño, digamos 0,000001; lo que hace es mostrar que si existe una hipótesis alternativa que pueda explicar la ocurrencia del resultado con una probabilidad razonablemente mayor, digamos 0,05, nos sentiremos mucho más inclinados a considerar que la hipótesis inicial no es cierta». Además, quisiera darte un consejo, que no es por cierto nada nuevo, sino que nos remite a la prédica más ortodoxa de la inferencia estadística clásica: nunca está justificado que te precipites a una empresa de verificación de hipótesis, sin tener una hipótesis alternativa fija y plausible que estés en condiciones de fundamentar desde el punto de vista de tu ciencia particular (la clínica, la epidemiología, la fisiología o la genética...). Más aún, la comunidad científica, comenzando por los comités editoriales de las revistas, tendrían derecho a exigir de los investigadores una fundamentación que avale la plausibilidad de la hipótesis en juego. A tenor con lo que hemos discutido, esta exigencia tendría el efecto de tornar en consenso común lo que en el mejor de los casos es una simple creencia individual y, desgraciadamente en muchas ocasiones, sólo una cacería fortuita de significación estadística.

Como diría Cnemon: «... Y ésta es, más o menos, la verdad sobre el asunto».

Epílogo

Con la anuencia entusiasta de Simplicio y la venia tolerante de Normancio y Lipidio, el relator querría poner fin a este primer contrapunto citando un ejemplo de Small y Schor (1983)¹³, tomado del *Annals of Internal Medicine*.

Veinte jugadores de cuadro de ligas menores de béisbol fueron asignados aleatoriamente a 2 grupos (de 10 cada uno) de entrenamiento, uno de los cuales aplicaba el procedimiento de rutina y el otro impartía además instrucciones para el control del estrés durante el partido.

Al concluir la temporada oficial los jugadores del primer grupo mejoraron su porcentaje de «fildeo» (lances sin error entre total de lances) en un promedio de 0,016, con una desviación estándar de 0,006. Los del grupo 2, aunque sufrieron la baja de un jugador que no pudo adaptarse a los rigores del entrenamiento anti-estrés, experimentaron una mejoría promedio de 0,025 con una desviación estándar de 0,0064.

Un estadístico clásico calculó el estadígrafo

$$t = (0,025 - 0,016) / (0,0062 [1/10 + 1/9])^{1/2}$$

y obtuvo un valor de 3,16, superior al valor crítico de 2,11 correspondiente a 17 grados de libertad y a una significación del 5%.

El bayesiano, por su parte, sin conocimiento previo alguno sobre el comportamiento de la variable en cuestión, utilizó una distribución a priori no informativa y obtuvo la variable *X* (diferencia en el cambio de la habilidad de fildeo con una y otra técnica de entrenamiento) que tiene una distribución a posteriori tal que

$$(X - [0,025 - 0,016]) / (0,0062 [1/10 + 1/90])^{1/2} = -3,16 + X/0,003,$$

se distribuye *t* con 17 grados de libertad. Esto significa que hay una probabilidad de 0,95 de que la diferencia en el cambio caiga en el intervalo 0,003 a 0,015 (que correspondería a una diferencia estadísticamente significativa puesto que no contiene el valor de la nulidad).

Ante el entusiasmo que despertaron los resultados, avalados por el juicio coincidente de los expertos, el entrenador de tercera apuntó que le preocupaba recomendar un método que estimulaba un 10% de deserción, y añadió –con cierto toque de cinismo– que una diferencia promedio de 0,009 se traducía sólo en 4,32 errores menos por jugador durante la campaña (tomando en cuenta que el número promedio de lances en una temporada es 480) y, dado que aproximadamente sólo uno de cada 20 errores tiene una influencia determinante en el resultado final del partido, todo lo que el entrenamiento conseguiría sería una victoria más en la temporada.

Conclusión

En este diálogo se esbozan los rasgos más generales de lo que constituye el fundamento epistemológico del enfoque bayesiano. Se exponen las razones de la inviabilidad inmediata de este enfoque, al tiempo que se sugieren las vías por las que habrá de convertirse en el modo natural de concebir el análisis estadístico con fines inferenciales. Se sugiere que mientras el modo de pensar bayesiano desplaza a las metodologías más clásicas basadas en las pruebas de significación (Fisher) y en las pruebas de hipótesis (Neumann-Pearson), conviene al menos atenerse a la práctica de recurrir a los métodos habituales de contrastación de hipótesis, sólo cuando puedan garantizarse bases de consen-

so para un alto grado de plausibilidad teórica de la hipótesis que se desea verificar.

REFERENCIAS BIBLIOGRÁFICAS

1. Feyerabend PK. Diálogos sobre el conocimiento (Colección Teorema). Madrid: Ediciones Cátedra, 1991.
2. Feyerabend PK. Diálogo sobre el método (Colección Teorema). Madrid: Ediciones Cátedra, 1990.
3. Salsburg D. Hypothesis versus significance testing for controlled clinical trials: A dialogue. *Stat in Med* 1990; 9: 201-211.
4. Fleiss JL. Significance tests have a role in epidemiologic research: reactions to AM Walker (Different Views). *Am J Pub Health* 1986; 76: 559-560.
5. Fleiss JL. Confidence intervals vs. significance tests: quantitative interpretations [carta]. *Am J Public Health* 1986; 76: 587.
6. Poole Ch. Beyond the confidence interval [Different Views]. *Am J Public Health* 1987; 77: 195-199.
7. Boen JR. What statistics can and cannot do. *Transfusion* 1993; 33: 940-943.
8. Parzen E. Modern probability theory and its applications. Nueva York: John Wiley, 1960; 119.
9. Light RJ. Accumulating evidence: using meta-analysis to carry out research reviews in Pediatrics. *Pediatrics* 1986; 78: 1.147-1.154.
10. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987; 9: 2-31.
11. Efron B. Why isn't everyone a Bayesian? *The American Statistician* 1986; 40: 1-5.
12. Hacking I. The logic of statistical inference. Cambridge, Cambridge University Press, 1965; 83.
13. Small RS, Schor SS. Bayesian and non-Bayesian methods of inference. *Ann Intern Med* 1983; 99: 857-859.