

Instituto Superior de Ciencias Médicas de La Habana

APUNTES SOBRE SUBJETIVIDAD Y ESTADÍSTICA EN LA INVESTIGACIÓN EN SALUD

Luis Carlos Silva Ayçaguer¹ y Alina Benavides Rodríguez²

RESUMEN: La tarea más importante de la estadística es proporcionar alternativas cuantitativas objetivas al mero juicio personal, de manera que nos proteja de la subjetividad y los sesgos en el proceso de obtención de nuevos conocimientos. Las pruebas de significación se consideran la expresión más encumbrada de ese afán de objetividad, pues se cree que pueden generar conclusiones independientemente de las personas que las emplean. Este trabajo procura fundamentar que esta es una falsa ilusión y que, en realidad, este enfoque tiene serias limitaciones conceptuales y prácticas que le impiden cumplir cabalmente la tarea que le dio origen, las cuales se tornan cada día más difíciles de soslayar. Lo cierto es que la subjetividad no es enteramente evitable, por lo que resulta mucho más útil contar con un método que combine las evidencias subjetivamente acumuladas, con la información objetiva obtenida de un experimento en particular. Los métodos bayesianos ofrecen esta posibilidad.

DeCS: ESTADÍSTICAS DE SALUD; TEOREMA DE BAYES.

INTRODUCCIÓN

La estadística ha sufrido un proceso que pudiéramos llamar de tránsito: de la marginalidad a la cima. Si bien 50 años atrás prácticamente no se utilizaba en las investigaciones médicas, en estos momentos muchos investigadores del ámbito sociológico, salubrista, epidemiológico o clínico sienten que sus trabajos no tienen suficiente rigor científico si no vienen avalados por al menos un proceder estadístico.¹ Por su carácter mecanicista, tal doctrina es insostenible, aunque sin duda esta disciplina puede muchas veces ocupar un importante espacio como recurso instrumental para operar en el mundo de la investigación médico-social.

Dicho de manera general, la tarea más importante de la estadística es proporcionar alternativas cuantitativas objetivas al mero juicio personal, que permitan una interpretación adecuada de los datos producidos por los estudios.² Este recurso no demoró en resultar extremadamente útil para los editores de revistas y responsables administrativos, pues a principios de siglo la ausencia de una herramienta que

aquilatara cuantitativamente el significado de los hallazgos propició que las anécdotas clínicas poblaran las revistas médicas. Se hacía necesario, por tanto, usar procedimientos que cuantificaran el peso probatorio de los resultados y que complementaran los razonamientos verbales, de modo que los protegiera de la subjetividad (Silva LC. La crisis de las pruebas de significación y la alternativa bayesiana. Memorias del XI Congreso de la Sociedad Gallega de Estadística e Investigación Operativa, 1999. Santiago de Compostela).

Lo cierto es que, aunque la objetividad es un deseo natural y legítimo, lamentablemente resulta inalcanzable en estado puro. La estadística no puede resolver este conflicto, pues todo proceso inferencial, incluso cuando se lleva adelante con el concurso de esta disciplina, tendrá siempre un componente subjetivo. Si bien las técnicas estadísticas pueden ser muy útiles, suelen generar una ilusión que es pernicioso en la medida que inyectan una convicción que conspira contra la obligación de examinar la realidad a través de un pensamiento integral. Nuestro propósito es desarrollar algunas ideas en la línea de fundamentarlo.

¹ Doctor en Ciencias. Investigador Titular, Profesor Auxiliar. Vicerrectoría de Investigación y Posgrado.

² Especialista de I Grado en Bioestadística. Hospital Universitario "Arnaldo Milián Castro", Villa Clara.

INEVITABILIDAD DE LO SUBJETIVO

En la investigación médico-social el componente subjetivo es ineludible, y en cierto sentido conveniente. Se supone que, al aplicar cierto procedimiento estadístico a un conjunto de datos, lo que se procura es que el análisis gane en objetividad; es decir, que los puntos de vista del investigador no puedan modificar sustancialmente las conclusiones, pero la verdad es que los métodos estadísticos están lejos de garantizar automáticamente tal *desideratum*.

Todos sabemos que la forma de operacionalizar las variables, los puntos de corte que se eligen, el nivel de significación empleado, las escalas de medición adoptadas, las pruebas de significación utilizadas, son solo algunos ejemplos de la larga lista de instrumentos estadísticos que no hay más remedio que elegir según un punto de vista que varía entre investigadores. Tal carencia de pautas uniformes es especialmente acusada en el punto culminante del proceso: a la hora de realizar inferencias una vez examinados los resultados. De hecho, cuando un paquete estadístico concluye su tarea, muchos investigadores creen que también ha terminado la suya. Se trata de un error esencial: en ese punto comienza la zona más importante de la tarea del investigador, quien no podrá eludir dar una impronta subjetiva a su análisis.

La creencia en la capacidad de la estadística para inyectar objetividad (e incluso, para garantizarla) tiene sus raíces en los procedimientos de pruebas de significación, ampliamente usados en la investigación médico-social. Sin embargo, como señalan *Berger* y *Berry*³ en un revelador artículo publicado hace ya algunos años en una prominente revista especializada, "...el uso común de la estadística parece haberse fosilizado, principalmente debido a la visión de que la estadística clásica es la forma objetiva de analizar datos".

LA PRUEBA DE SIGNIFICACIÓN ¿PARADIGMA DE OBJETIVIDAD?

Sin dudas, su aparición en escena representó un cambio importante con respecto a los métodos anteriores, ya que se trata de un proceder que esencialmente dictaba las acciones del investigador. De hecho, se convirtió en una estrategia ampliamente promovida y aceptada en el ámbito investigativo; tal acogida se debió, probablemente a que, como señala *Goudman*,⁴ tanto para los investigadores como para los editores de revistas y responsables administrativos, resulta muy atractivo contar con procedimientos cuantitativos capaces de generar conclusiones independientemente de las personas que realizan el estudio.

Sin embargo, en realidad la prueba de significación actual procede de la unión de dos perspectivas originalmente divergentes, lo cual ha dado lugar a un procedimiento con serias limitaciones conceptuales y prácticas. Como

apunta el importante estadístico y epistemólogo norteamericano *Richard Royall*,² ni la literatura general ni los programas docentes de estadística informan de sus contradicciones. Tampoco se da cuenta del intenso debate que estas han animado durante casi 70 años entre muchos estadísticos que se cuestionan (o defienden) la "solidez" de las pruebas de significación.⁵⁻⁶

Este escamoteo ha contribuido a abonar en la mayoría de los profesionales sanitarios la errónea convicción de que sus resultados tendrán más rigor científico por el solo hecho de venir acompañados de "valores p".

El debate en torno a las pruebas de significación en sociología data de la década del 50, pero sus raíces se remontan a un artículo de *Margaret Hagood* publicado en 1941 bajo el título "Estadística para sociólogos".⁷ Desde entonces, muchas observaciones críticas se han venido acumulando, hasta conformar hoy un reclamo metodológico de tal magnitud que cada día se torna más difícil soslayar.

Entre las diversas objeciones que se le hacen al método,⁸ la que mejor viene al caso es que no toma en cuenta de manera formal en el modelo de análisis, la información anterior a los datos actuales, la que proviene de estudios previos, o de la experiencia empírica informalmente acumulada que siempre se tiene sobre el problema que se examina. Esto supone un vacío de opiniones, una orfandad total de información, que es irreal en la práctica.⁹

El precio que tenemos que pagar para disfrutar los beneficios de la "objetividad" inherente a estas pruebas es abandonar nuestra capacidad para juzgar presuntas verdades en un estudio individual. En la práctica esto significa que lo único que puede hacerse es comunicar si los resultados son estadísticamente significativos o no, y actuar de acuerdo con ese veredicto. Para algunos, se trata de un hecho profundamente anticientífico y contrario a la intuición. Ello explica que no sea infrecuente hallar trabajos en que el texto analítico que subsigue a la aplicación de la prueba parece desentenderse de las propias reglas que esta impone. Por eso no es extraño leer frases del tipo "aunque la diferencia no es significativa, nótese que...", o del tipo "no hemos hallado significación, pero con una muestra más grande...". A este último caso destinamos el siguiente párrafo.

Otras críticas están relacionadas con el procedimiento *per se* y cuestionan seriamente el carácter objetivo que se les atribuye. La endeblez más grave en este sentido es que, dada la naturaleza de los valores p, el rechazo o la aceptación de una hipótesis resulta ser, simplemente, un reflejo del tamaño de la muestra: si esta es suficientemente grande, siempre se rechazará la hipótesis nula. Esto nos coloca en una aparente paradoja: cuando operamos con una parte muy pequeña de la realidad (una muestra muy pequeña), entonces no podemos obtener conclusión alguna, como es lógico e intuitivo, lo cual conduce a que muchos investigadores, cuyos resultados no alcanzan la esperada significación estadística, proclamen que con un tamaño de muestra

mayor casi seguramente lo hubieran logrado. Lamentablemente, y esto es lo realmente grave, tienen razón; pero eso significa que tampoco se puede sacar nada en claro cuando se trabaja con una muestra muy grande, puesto que en tal caso el rechazo de la hipótesis nula queda virtualmente asegurado.¹⁰

Por otra parte, para ilustrar el mal uso que se hace del procedimiento, baste reparar en la confusión que se ha establecido entre significación estadística y significación clínica o social. A pesar de que muchos han argüido que la decisión no debe basarse exclusivamente en la significación estadística sino que ha de atenderse también a la significación sustantiva clínica o social, este error se continúa cometiendo. Según *Hagood*,⁷ en la prisa por aplicar las pruebas de significación, los investigadores olvidaron que “significación estadística” es simplemente la base para el rechazo de una hipótesis nula. En este sentido, hace más de 15 años *Feinstein*¹¹ apuntaba:

"Si la demanda crítica hubiera sido que la investigación produjese ambos tipos de significación (la que concierne al área estocástica y la que se vincula con los atributos cualitativos) entonces la alienación intelectual de hoy no hubiera ocurrido. Sin embargo, desafortunadamente la palabra *significación* fue reservada sólo en su connotación estocástica, y la palabra *estadística* le fue adjuntada para crear la *significación estadística* como paradigma de calidad e importancia en la investigación médica...".

Cohen,¹² en un artículo cuyo ingenioso título resume una patética realidad cotidiana, escribía cáusticamente:

"Después de cuatro décadas de duras críticas, el ritual de probar la significación de hipótesis nulas –decisiones dicotómicas mecánicas alrededor del sacralizado criterio de 0,05 –aún persiste. [...] ¿Cuál es el problema con las pruebas de significación? Bueno, que entre otras cosas, no nos dicen lo que queremos saber, y deseamos tanto saber que lo que queremos saber es esto, que por desesperación, sin embargo, creemos que lo hace".

La situación ha llegado a tal punto que en sus recomendaciones de octubre del 2001, el Comité Internacional de Directores de Revistas Médicas (Grupo de *Vancouver*)¹³ llega a establecer que: “Se evitará la dependencia exclusiva de las pruebas estadísticas de verificación de hipótesis, tal como el uso de los valores *p*, que no aportan ninguna información cuantitativa importante” (Comité Internacional de Directores de Revistas Médicas. Octubre, 2001. Requisitos uniformes de los manuscritos enviados a revistas biomédicas).

Pero si la significación estadística no aporta nada cuando no se involucra en el análisis a la significación clínica, entonces el componente subjetivo que se asocia inexorablemente a esta última, mediatiza automáticamente a la primera.

El procedimiento de prueba de significación actual supuso un cambio en materia de autoridad médica, que pasó de estar en manos de aquellos con conocimiento de las bases biológicas y sociales de la medicina a las de quienes

dominan los métodos cuantitativos, o hacia los resultados cuantitativos solamente, como si los números pudieran hablar por ellos mismos.¹⁴

Lo inquietante en el procedimiento clásico no radica en que la subjetividad desempeñe un papel en el proceso; sino que en lugar de reducir la influencia del componente subjetivo, las pruebas de hipótesis puedan servir en realidad solamente para ocultar su participación.

MÉTODOS BAYESIANOS: UNA NOTA FINAL

Lo cierto es que la subjetividad (algo bien diferente a la arbitrariedad o al capricho) es un fenómeno inevitable, especialmente en un marco de incertidumbre como el de las ciencias biológicas y sociales; siempre aparecerá en una u otra medida, e independientemente del recurso estadístico que se emplee. Por tanto, resultaría mucho más útil, además de ser consistente con las demandas de la intuición, contar con un método que combine las evidencias subjetivamente acumuladas con la información objetiva obtenida de un experimento en particular. Los métodos bayesianos ofrecen esta posibilidad.

Si bien la exposición de este recurso exige un nivel teórico que desborda el que consideramos oportuno para estas reflexiones, procede destacar que se trata de una aproximación metodológica que está a salvo de casi todas las impugnaciones que se le hacen a las pruebas de significación y que goza del atractivo de incorporar formal y explícitamente los conocimientos aportados por experiencias previas dentro del proceso analítico y los contempla, por tanto, en las conclusiones.¹⁵ Tiene sus propios conflictos, pero ahora interesa subrayar uno de sus méritos indiscutibles: asume sin ambages la existencia de un marco subjetivo ineludible en el análisis y, en lugar de actuar como si no existiera, obligan al investigador a expresar probabilísticamente sus puntos de vista apriorísticos y luego incorpora ese resultado en el análisis formal de los datos. Tal es el planteamiento esencial del ya citado trabajo de *Berger* y *Berry*³ cuando llaman a descartar la noción de que la estadística puede ser incontaminadamente objetiva, y a “abrazar la necesidad de la subjetividad a través del análisis bayesiano, el cual puede conducir a un análisis de datos más flexible, poderoso y comprensible”.

Según nuestro juicio, el proceso intelectual asociado a la inferencia bayesiana es mucho más coherente con el pensamiento natural del científico que el que se vincula con las pruebas clásicas de hipótesis. Estos procedimientos constituyen una tecnología emergente de procesamiento y análisis de la información para la que cabe esperar una presencia cada vez más intensa en el campo de la aplicación de la estadística a la investigación empírica contemporánea. Para tener una idea de en qué medida tal expectativa dista de ser un afiebrado deseo, basta detenerse en el elocuente resumen de sus aplicaciones recientemente publicado por *David Malakoff*¹⁵ en la prestigiosa revista *Science*.

Summary: The most important task of statistics is to provide the mere personal judgement with quantitative objective alternatives in order to protect us from subjectivity and the sesgos in the process of obtention of new knowledge. The significance tests are considered the highest expression of this eagerness of objectivity, since it is believed that they may generate conclusions independently of the persons using them. This paper intends to establish that it is a false illusion and that, in reality, this approach has serious conceptual and practical limitations preventing it from fulfilling precisely the task that gave origin to it and which are more difficult every day to be ignored. It is true that subjectivity cannot be completely avoided and that's why it is more useful to have a method combining the subjectively accumulated evidences with the objective information obtained from an experiment in particular. Bayes' methods offer this possibility.

Subject headings: **HEALTH STATISTICS; BAYES THEOREM.**

REFERENCIAS BIBLIOGRÁFICAS

1. Silva LC. Cultura estadística e investigaciones en el campo de la salud: una mirada critica. Madrid: Díaz de Santos; 1997.
2. Royall RM. Statistical evidence: a likelihood paradigm. Boca Raton: Chapman & Hall/CRC; 1997.
3. Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *Am Scientist* 1988;76:159-65.
4. Goodman SN. Toward evidence-based medical statistics (I): The p value fallacy". *Ann Int Med* 1999;130:995-1004.
5. Morrison DE, Henkel RE. The Significance test controversy –A Reader. Chicago: Aldine Publishing Company; 1970.
6. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychol Methods* 2000;5:241-301.
7. Hagood MJ. Statistics for sociologists. New York: Reynal and Hitchcock; 1941.
8. Silva LC. La alternativa bayesiana. *Brotos* 2000;1:1-4.
9. Silva LC, Benavides A. El enfoque bayesiano: otra manera de inferir. *Gaceta Sanit* 2001;15:341-6.
10. Benavides A, Silva LC. Contra la sumisión estadística: un apunte sobre las pruebas de significación. *Metas Enferm* 2000;3:35-40.
11. Feinstein AR. Clinical epidemiology: The architecture of clinical research. Philadelphia: W.B. Saunders Company; 1985.
12. Cohen J. The earth is round ($p < .05$). *Am Psychol* 1994;49:997-1003.
13. Silva LC. Hacia una cultura epidemiológica revitalizada. *Dimensión Hum* 1997;1(5):23-33.
14. Silva LC, Suárez P. ¿Qué es la inferencia bayesiana?. *JANO, Med Hum* 2000;58(1338):65-6.
15. Malakoff D. Bayes offers a 'new' way to make sense of numbers. *Science* 1999;286:1460-4.

Recibido: 8 de enero del 2003. Aprobado: 29 de enero del 2003.
Luis Carlos Silva Ayzaguer. Vicerrectoría de Investigación y Posgrado,
 Instituto Superior de Ciencias Médicas de La Habana. E-mail:
 lcsilva@infomed.sld.cu