



Estadística bayesiana: un nuevo paradigma para incorporar en la investigación clínica

Mario Enrique Rendón-Macías^{1,*}

¹ Unidad de Investigación en Análisis y Síntesis de la Evidencia, Coordinación de Investigación en Salud, Instituto Mexicano del Seguro Social. Universidad Panamericana, Escuela de Medicina, Ciudad de México.

En las últimas décadas, con el propósito de mejorar la inferencia estadística, en el mundo se está retomando el análisis bayesiano. Si bien, estamos conscientes de que pasará tiempo para su aceptación e incorporación, muy probablemente no sustituirá a la estadística clásica, es decir, la estadística frecuentista. De esta forma, consideramos importante que los investigadores y lectores de la literatura médico-científica la conozcan para que puedan comprenderla mejor.¹

En general, la estadística busca simplificar o resumir información de una investigación para entender, analizar e inferir conclusiones útiles al tomar decisiones. Sus dos objetivos principales son la estimación de parámetros (ej. medias, proporciones, porcentajes, tasa, etc.) y ejecución de pruebas de significancia estadística de la hipótesis nula (PSEHN).¹⁻³

Bajo la estadística clásica o frecuentista, cuando se estima un parámetro es necesario agregar su precisión mediante el intervalo de confianza al 95% (IC_{95%}). Por otro lado, la PSEHN permite determinar la plausibilidad de los datos obtenidos en un estudio, para rechazarla o no, bajo el criterio de un nivel crítico pre establecido, siendo el más común $p < 0.05$.²

A pesar de su alta aceptación y difusión, la estadística frecuentista ha sido criticada por lo siguiente: 1. En la determinación y precisión de un parámetro. Cualquier valor que esté dentro del IC_{95%} tiene la misma probabilidad de certeza; además, este intervalo

pudiera no incluir al verdadero parámetro.^{2,4} 2. Bajo la PSEHN, sólo se contrasta la hipótesis nula, la cual por definición es cierta. Por tanto, no es posible confirmar la validez de la hipótesis alterna, la cual es de interés del investigador.^{4,5} 3. En un intento por mejorar la interpretación de los resultados de una investigación, se ha sugerido la presentación del tamaño del efecto considerando umbrales clínicos de decisión.^{6,7} 4. El análisis estadístico frecuentista no incorpora la evidencia previa a un estudio, es decir, los cálculos de las pruebas estadísticas sólo utilizan los datos obtenidos del estudio en cuestión.^{3,5}

La estadística bayesiana (*Tabla 1*), aunque surgida un siglo antes de la frecuentista, ha resurgido en años recientes.⁸ Contrario a la frecuentista, el cálculo es más complicado, pero su crecimiento se debe al desarrollo de nueva tecnología (mejores procesadores y software).⁵⁻¹⁰ Es conveniente señalar que este tipo de estadística también permite estimar parámetros y contrastar hipótesis de investigación, pero tiene fundamentos distintos a la frecuentista. El primero consiste en que para la estadística bayesiana se asume que realizamos estudios con base en observaciones previas o razonamientos teóricos, los cuales son traducidos en probabilidades de credibilidad de la verdad o veracidad. El investigador realiza "solo" su estudio para ganar información y, posiblemente, modificar su nivel de certidumbre a su conocimiento previo. Los resultados del

* Correspondencia: MERM, drmariorendon@gmail.com

Conflictos de intereses: El autor declara que no tiene.

Citar como: Rendón-Macías ME. Estadística bayesiana: un nuevo paradigma para incorporar en la investigación clínica. Rev Mex Pediatr 2019; 86(2):43-46.

[Bayesian statistics: a new paradigm to incorporate in clinical research]

estudio actualizan su credibilidad, lo que se denomina probabilidad posterior.¹¹⁻¹³

Para la estimación de un efecto (por ejemplo, diferencia de un promedio entre grupos), los investigadores antes de realizar su estudio determinan las probabilidades de encontrar cada tamaño del efecto (es decir, el parámetro). Estas probabilidades se resumen en una distribución previa (para más detalles, consultar la referencias). Despues, tras la realización de la investigación y con base en los resultados se calculan nuevamente las probabilidades de parámetros posibles y se estima el más probable. Los parámetros más probables son resumidos en “el intervalo de credibilidad” (o intervalo de mayor densidad) al 95%. Este intervalo sí se interpreta como las cifras más posibles en el 95% de las veces que se haría un estudio y el valor puntual encontrado es “el más probable”. Por eso, aunque el intervalo de credibilidad al 95% incluya al valor nulo, si éste no es el más probable podemos hablar de un efecto con un grado de credibilidad. Si el intervalo no incluye al valor nulo, la credibilidad es al menos del 95% (*Tabla 1*). Si el intervalo está más allá de un rango equivalente a la nulidad (ROPE), la evidencia es muy fuerte hacia la hipótesis alterna; mientras que si éste se encuentra dentro de ese rango de equivalencia, se apoya fuertemente a la hipótesis nula.¹⁴⁻¹⁸

La prueba de hipótesis bayesiana no se basa en el rechazo de una hipótesis nula, sino en poder contrastar dos o más. En general, los estudios sólo contrastan dos hipótesis: la nula o no-efecto (H_0) contra la alterna o efecto (H_1). Una vez realizado el estudio se estima la probabilidad de los datos bajo la hipótesis nula y, la misma, bajo la alterna. La relación entre estas dos hipótesis se resume en el “factor Bayes o FB”.

Con el factor Bayes, existen dos formas: FB10 o razón $H_1:H_0$ y FB01 o razón $H_0:H_1$. La primera nos dice cuántas veces son más compatibles los datos con la H_1 contra H_0 y la segunda la relación inversa. La información del FB de un estudio permite realmente hablar de apoyo a una hipótesis (tanto nula como la alterna). Entre más alejado este el FB del valor 1 (valor de no predominio de alguna hipótesis, o escasa información para tomar una decisión) una de las dos hipótesis tendrá mayor credibilidad (*Figura 1*).¹⁹⁻²²

La decisión final sobre la validez de una hipótesis se basa en la probabilidad final del estimado actualizado con la conjunción de su probabilidad previa, por la aportación de los datos del estudio (verosimilitud de los datos). Así, entre mayor evidencia exista en los estudios previos a favor de una hipótesis, los datos aportados en un estudio podrían modificar poco las conclusiones. Por el contrario, ante escasa evidencia o contraria a la previa, un estudio debe aportar datos muy contundentes

Tabla 1: Correspondencia entre significancia estadística entre la estadística frecuentista o clásica contra la Bayesiana.

	Estadística frecuentista	Estadística bayesiana
Criterio de prueba de hipótesis	Solo contrasta la hipótesis nula, valor de “p” Criterio dicotómico (se rechaza o no)	Contrasta dos hipótesis (pueden ser más) H_n nula contra H_a , El factor Bayes da sustento a la más probable o escasa evidencia para concluir
Estimación del valor de parámetro y la incertidumbre	Intervalo de confianza del efecto estimado bajo un nivel pre establecido (valor-p) Todos los valores dentro del intervalo tienen la misma probabilidad de ser el estimado paramétrico	Distribución posterior del efecto, intervalo de credibilidad o de alta densidad Muestra los valores posibles del parámetro y señala el valor más probable
Interpretación de resultados	$p <$ nivel crítico: se rechaza hipótesis nula $p \geq$ nivel crítico: no se rechaza hipótesis nula Intervalo de confianza al 95% no incluye al valor nulo: significancia estadística Intervalo de confianza al 95% incluye al valor nulo: no significancia estadística $IC_{95\%}$ estrecho: alta precisión $IC_{95\%}$ amplio: baja precisión	Interpretación factor Bayes (<i>Figura 1</i>) Intervalo de credibilidad al 95% posterior a la actualización de los datos no incluye al valor nulo o equivalente al nulo: 95% de certeza del efecto Intervalo de credibilidad al 95% posterior a la actualización de los datos que incluye al valor nulo o equivalente al nulo: estimar el % de certeza del efecto por arriba de ese valor

	FB_{01} Hipótesis nula	Valor	FB_{10} Hipótesis alterna
A favor	Muy fuerte	>30	Muy fuerte
	Fuerte	10-30	Fuerte
	Moderada	3.1-10	Moderada
	Anecdótica	1.1-3	Anecdótica
En contra	No evidencia	1	No evidencia
	Anecdótica	0.3-0.9	Anecdótica
	Moderada	0.29-0.1	Moderada
	Fuerte	0.09-0.03	Fuerte
	Muy fuerte	< 0.03	Muy fuerte

Figura 1: Interpretación de factor de Bayes (FB). En la FB_{01} se contrasta la hipótesis nula sobre la alterna, en la FB_{10} la alterna sobre la nula.

para cambiar nuestra impresión inicial. Es importante señalar que la selección de cuál información previa usar para el cálculo inicial, ha sido una objeción al uso de la estadística bayesiana. Esto último se puede superar con un análisis de sensibilidad, considerando información previa, ya sea ausente (no informativa), optimista (a favor de una hipótesis de estudio) o escéptica (en contra de la hipótesis de estudio).²³⁻²⁶

Finalmente, destacamos que en el presente y un futuro inmediato, la estadística bayesiana se podrá incorporar en dos áreas donde la estadística frecuentista tiene limitaciones: a) en investigaciones donde el tamaño de muestra es pequeño, como ocurre en muchos estudios clínicos y, b) al realizar análisis complejos pero, sin la problemática de cumplir los supuestos de normalidad.^{5,11,19,25,27}

REFERENCIAS

- Hackenberger BK. Bayes or not Bayes, is this the question? *Croat Med J*. 2019; 60: 50-52.
- Curran-Everett D. Exploration in statistics: hypothesis test and P values. *Adv Physiol Educ*. 2009; 33: 81-86.
- Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole Ch, Goodman SN et al. Statistical test, p values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016; 31: 337-350.
- Goodman SB. Aligning statistical and scientific reasoning. Misunderstanding and misuse of statistical significance impede science. *Science*. 2016; 352(6290): 1180-1181. doi: 10.1126/science.aaf5406.
- Dienes Z, McLatchie N. Four reasons to prefer Bayesian analyses over significance testing. *Psychon Bull Rev*. 2017. doi: 10.3758/s13423-017-1266-z
- Martínez-Ezquerro JD, Riojas-Garza A, Rendón-Macías ME. Clinical significance vs statistical significance. How to interpret the confidence interval at 95%. *Rev Alerg Mex*. 2017; 64(2): 220-227.
- Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers EJ. The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev*. 2016; 23: 103-123. doi: 10.3758/s13423-015-0947-8.
- Greenland S, Poole CH. Living with P values. Resurrecting a Bayesian perspective on frequentist statistics. *Epidemiology*. 2013; 34: 62-68.
- Zangiacomi-Martínez E, Achcar JA. Trends in epidemiology in the 21st century: time to adopt Bayesian methods. *Sau Pub Rio Janeiro*. 2014; 30(4): 703-714. doi: 10.1590/0102-311X00144013.
- Asby D. Bayesian statistics in medicine: a 25 years review. *Stat Med*. 2006; 25: 3589-3631.
- Austin PC, Brunner LJ, Hux JE. Bayes watch: an overview of Bayesian statistics. *J Eval Clin Pract*. 2002; 8(2): 277-286.
- Gurrin LC, Kurinczuk JJ, Burton PR. Bayesian statistics in medical research: an intuitive alternative to conventional data analysis. *J Eval Clin Pract*. 2000; 6(2): 193-204.
- Bittl JA, He Y. Bayesian analysis. A practical approach to interpret clinical trials and clinical practice guidelines. *Circ Cardiovasc Qual Outcomes*. 2007; 10.e003563. doi: 10.1161/CIRCOUTCOMES.117.003536.
- Etz A, Gronau QF, Dablander F, Edelsbrunner PA, Baribault B. How to become a Bayesian in eight easy steps: an annotated reading list. *Psychon Bull Rev*. 2017; 28. doi: 10.3758/s13423-017-1317-5.
- McCulloch TJ. Bayesian statistics: how to quantify uncertainty. *Anaest Inten Care*. 2011; 39: 1001-1003.
- Haskins R, Osmotherly PG, Tuyl F, Rivett DA. Uncertainty in clinical prediction rules: the value of credible intervals. *J Orthop & Sport Therap*. 2014; 44(2): 85-91.
- Wagenmakers EJ, Love J, Marsma M et al. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychon Bull Rev*. 2017. doi: 10.3758/s13423-017-1323-7.

18. Matthews RA. Beyond "significance": principles and practice of the analysis of credibility. *R Soc open Sci.* 2017; 5: 171047. doi: 10.1098/rsos.171047.
19. Kruschke JK, Liddell TM. The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from Bayesian perspective. *Psychon Bull Rev.* 2017. doi: 10.3758/s13423-016-1221-4.
20. Shin JJ, Zurakowski D. Null hypotheses, interval estimation, and Bayesian analysis. *Otolaryngology-Head Neck Surg.* 2017; 157(6): 919-920.
21. Rendón-Macías ME, Riojas-Garza A, Contreras-Estrada D, Martínez-Ezquerro JD. Bayesian analysis. Basic and practical concepts for interpretation and use. *Rev Alerg Mex.* 2018; 65(3): 205-218.
22. Weiss RE. Bayesian methods for data analysis. *Am J Ophthalmol.* 2010; 149(2): 187. doi: 10.1016/j.ajo.2009.11.011.
23. Jeon M, De Boeck P. Decision qualities of Bayes factor and p value-based hypothesis testing. *Am Psych Meth.* 2017; 22(2): 340-360.
24. Hoekstra R, Moden R, van Ravenzwaaij D, Wagenmakers EJ. Bayesian reanalysis of null results reported in medicine: Strong yet variable evidence for the absence of treatment effects. *PLoS ONE.* 2018; 13(4): e0195474. doi: 10.1371/journal.pone.0195474.
25. Wilcox RR, Serang S. Hypothesis testing, p values, confidential intervals, measures of effect size, and bayesian methods in light of modern robust techniques. *Educ & Psychol Measement.* 2017; 77(4): 673-689.
26. Quintana DS, Williams DR. Bayesian alternatives for common null-hypothesis tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry.* 2018; 18: 178. doi: 10.1186/s12888-1761-4.
27. López-Puga J, Krzywinski M, Altman N. Bayes theorem. Incorporate new evidence to update prior information. *Nat Meth.* 2015; 12(4): 277-278.