

Carta a la editora

Medwave 2014;14(1):e5894 doi: 10.5867/medwave.2014.01.5894

Intervalos de confianza y valores p

Confidence intervals and p values

Autor: Luis Carlos Silva Ayçaguer⁽¹⁾

Filiación:

⁽¹⁾Centro Nacional de Información de Ciencias Médicas, La Habana, Cuba

E-mail: lcsilva@infomed.sld.cu

Citación: Silva LC. Confidence intervals and p values. *Medwave* 2014;14(1):e5894 doi: 10.5867/medwave.2014.01.5894

Fecha de publicación: 20/2/2014

Señora editora:

He leído con interés el artículo titulado "Estadística para aterrorizados: interpretando intervalos de confianza y valores p", recientemente publicado en su revista¹. El texto aborda un tema de gran importancia para el trabajo cotidiano de los investigadores y por eso mismo concita atención crítica. En lugar de valorarlo detalladamente, opto por poner a su consideración algunas reflexiones relacionadas con tres afirmaciones allí realizadas. No aludo a errores graves. De hecho, son aseveraciones que se hacen con bastante frecuencia. Pero a mi juicio, merecen ser matizadas en beneficio del rigor y de los lectores, quienes podrán así conocer de otra perspectiva.

1. En el artículo se afirma: "*Como es casi imposible trabajar con una población completa, nunca se sabrá si la hipótesis nula es cierta con certeza absoluta*".

La realidad es otra. Salvo contadísimas excepciones, ocurre lo contrario: virtualmente siempre se sabe que la hipótesis nula es falsa. Esta es una de las endebles más importantes de la teoría de las pruebas de hipótesis: la de estar concebida con la exclusiva finalidad de responder a una pregunta para la cual ya se tiene la respuesta.

De hace ya media centuria data la siguiente observación, insertada en un artículo del influyente *Psychological Bulletin*: "*Es un hecho objetivo que casi nunca hay buenas razones para esperar que la hipótesis nula sea verdadera. ¿Por qué razón la media de los resultados de cierta prueba habría de ser exactamente igual al este que al oeste del río Mississippi? ¿Por qué deberíamos esperar que un coeficiente de correlación poblacional sea igual a 0,00? ¿Por qué esperar que la razón mujeres/hombres sea exactamente 50:50 en una comunidad dada? o ¿Por qué dos drogas habrán de producir exactamente el mismo efecto? Una mirada a cualquier conjunto de estadísticas que incluyan totales poblacionales confirmará de inmediato que la nulidad se presenta muy raramente en la naturaleza*"². Para poner un ejemplo concreto, similar al empleado con fines didácticos en un artículo reciente de Medwave³, consideremos que se discute si existe relación entre emplear cinturón de seguridad y el nivel

socioeconómico del conductor, reducido éste a dos niveles (alto y bajo). Si el OR correspondiente a la relación entre emplear cinturón de seguridad y un alto nivel socioeconómico fuera exactamente igual a 1, estaríamos presenciando un hecho francamente asombroso, pues la asociación pudiera ser muy pequeña, pero no es razonable imaginar que no exista absolutamente ninguna relación entre ambas condiciones. Resultaría insólito que la asociación fuera nula.

Lo que con toda seguridad va a ocurrir es que, si la muestra empleada es suficientemente grande, se podrá declarar que el valor de OR difiere de la unidad de manera estadísticamente significativa. Si dos hormigas nos parecen idénticas es porque no las hemos mirado con suficiente detención; no hay dos hormigas exactamente iguales. No hace falta una lupa suficientemente grande para saberlo. Para profundizar sobre este problema, sugiero al lector la lectura del texto "Una fábula significativa" (<http://lcsilva.sbhac.net/Otros/Aplicacionest/Trayectos/trayectos.htm>), donde se hace un examen sumamente detallado de este asunto.

2. Refiriéndose a la confiabilidad con que suelen construirse los intervalos de confianza, los autores expresan textualmente: "*Esta probabilidad ha sido fijada por consenso en un 95% en base a supuestos de normalidad, pero rangos entre el 90% y 99% son comúnmente utilizados en la literatura científica*".

El "pero" posterior a la coma, autoriza a pensar que en la frase final se está aludiendo a otros valores diferentes del 95%; y cuando se usa la palabra "rangos", entiendo que con ese término se quiere aludir a "valores". Supongo, en fin, que los autores han querido expresar esto otro: "*Esta probabilidad ha sido fijada por consenso en un 95% en base a supuestos de normalidad, pero otros valores entre el 90% y 99% son comúnmente utilizados en la literatura científica*".

Lo del consenso es ciertamente discutible. A mi juicio sería más correcto decir que ha sido fijada por inercia con la propuesta de Ronald Fisher hace nueve décadas⁴. Y la inercia es tan poderosa que resulta falso que otros valores, tales como 90% y 99%, sean "comúnmente" utilizados (mucho menos comunes son valores

intermedios que no sean 95%, los cuales, en realidad, no se utilizan virtualmente jamás). Una estimación rápida a partir de las entradas en Google Scholar - circunscribiéndonos al mundo académico- arroja que, de 789 mil ocasiones en que se alude a un *confidence interval*, 733 mil conciernen a intervalos calculados con una confiabilidad del 95%.

3. Finalmente, en el trabajo se comunica que: *Hacer pruebas de hipótesis (valores p) o estimar (intervalos de confianza) son técnicas validadas que contribuyen a la precisión de los efectos clínicos de cualquier investigación original.*

La oración resulta confusa. No se entiende qué significa que estas técnicas contribuyan "a la precisión de los efectos clínicos". Los efectos clínicos de alguna intervención se pueden aquilatar o valorar mediante esas técnicas, pero ellas no le confieren ni más ni menos precisión a tales efectos. De hecho, carece de sentido hablar de la "precisión de los efectos clínicos". Los efectos clínicos son los que son; las que pueden ser más o menos precisas son las estimaciones de la magnitud que dichos efectos tengan.

Pero, más allá de este desliz sintáctico, la aseveración de que tanto el empleo de valores p como la estimación de los efectos mediante intervalos de confianza son técnicas *validadas* para valorar los efectos clínicos, resulta difícil de aceptar. El número de artículos científicos que, a lo largo de los últimos decenios, cuestionan el uso de los valores p con este fin se cuenta por decenas o incluso cientos. La intensidad con que se emplean los valores p -con no poca frecuencia de manera ceremonial o ritual⁵ - no legitima que los consideremos *validados*. En lugar de extenderme en la fundamentación de esta afirmación, creo que basta remitirnos a lo que sugiere al respecto el llamado "Grupo de Vancouver", cuyas recomendaciones son explícitamente asumidas como patrón en la revista que Ud. dirige. Ellas vienen a decir que resulta inaceptable que un trabajo se base solamente en los valores p. Según esa fuente, se puede contemporizar con ellos, pero siempre que se acompañen de intervalos de confianza; en cambio, se puede prescindir de ellos si en su lugar se emplean solo dichos intervalos.

Debe decirse que tal regla, aunque cada vez menos, suele ser violentada por parte de autores y editores⁶, pero la mencionada advertencia es clara e inequívoca. La cita

textual en inglés es: "When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). **Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size** (énfasis añadido por mí, LCS)". Esta recomendación, por cierto, ha sido reivindicada como una "excelente declaración" en la guía para la comunicación de resultados estadísticos que acaba de ver la luz⁸, elaborada con la colaboración del estadístico histórico del *BMJ*, Douglas Altman.

Permítame concluir con una felicitación al trabajo que viene realizando su revista, la cual, además de contener interesantes e imprescindibles trabajos de gran actualidad técnica y política, ofrece un espacio de reflexión e intercambio de alto valor.

Referencias

1. Madrid E, Martínez F. Estadística para aterrorizados: interpretando intervalos de confianza y valores p. *Medwave* 2014;14(1):5892. | [CrossRef](#) |
2. Bakan D. The test of significance in psychological research. *Psychological Bulletin*. 1966;66:423-437. | [CrossRef](#) |
3. Quevedo F. La prueba de ji-cuadrado. *Medwave* 2011;11(12):e5266 | [CrossRef](#) |
4. Fisher RA. *Statistical methods for research workers*. Edinburgh: Oliver & Boyd, 1925. | [Link](#) |
5. Silva LC. Los límites de las pruebas de significación estadística y los valores p. *AMF*. 2008;4(2):62-63. | [Link](#) |
6. Silva-Ayçaguer LC, Suárez-Gil P, Fernández-Somoano A. The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. *BMC Med Res Methodol*. 2010 May 19;10:44. | [CrossRef](#) | [PubMed](#) | [PMC](#) |
7. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication. *ICMJE*, 2011. icmje.org [on line] | [Link](#) |
8. Lang T, Altman D. Basic statistical reporting for articles published in clinical medical journals: the SAMPL guidelines. En: *Science Editors' Handbook*. EASE, 2013.

Correspondencia a:

Calle 27 e/ M y N #110.
Vedado, 10400
Ciudad de la Habana
Cuba



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.