

Tópicos en investigación clínica y epidemiológica

# DEFINICIÓN E INTERPRETACIÓN ADECUADA DEL $p$ -valor E INTERVALOS DE CONFIANZA EN INVESTIGACIÓN CLÍNICA

Fredy Orlando Mendivelso<sup>1</sup>, Ingrid Milena Rodríguez<sup>2</sup>

<sup>1</sup> MD y Cirujano. Esp. MPH. MSc. FETP. Clínica Reina Sofía. Excelencia Clínica. Vicepresidencia Global de Salud-Keralty

<sup>2</sup> MV. Epidemióloga. MSc. FETP. Fundación Universitaria Sanitas

## RESUMEN

El  $p$ -valor y los intervalos de confianza se utilizan ampliamente en investigación clínica; sin embargo, no siempre es clara su definición y aún más, su uso por parte de los investigadores. La interpretación equivocada de estos dos tópicos estadísticos ha generado la falsa sensación en la comunidad médica de encontrar resultados con relevancia clínica siempre y cuando estos sean matizados por reportar valores “estadísticamente significativos” según el  $p$ -valor. De forma similar ocurre con el intervalos de confianza 95% porque no siempre se ha reportado e interpretado con una relación al estimador del efecto en el estudio. Nuestro propósito fue presentar la forma más clara posible, la definición, los usos y la correcta interpretación del  $p$ -valor e intervalos de confianza 95%

**Palabras clave:** Intervalo de confianza; Pruebas de Hipótesis; Bioestadística, Interpretación Estadística de Datos; Protocolos Clínicos

## DEFINITION AND APPROPRIATE INTERPRETATION OF P-VALUE AND CONFIDENCE INTERVALS IN CLINICAL RESEARCH

## ABSTRACT

The  $p$ -value and confidence intervals are widely used in clinical research; however, is not always clear its definition and even more, its use by researchers. The wrong interpretation of these two statistical topics has generated the false sensation in the medical community of finding results with clinical relevance as long as they are nuanced as “statistically significant” values according to the  $p$ -value. In a similar way, it occurs with confidence intervals 95% because it has not always been reported and interpreted with

Recibido: 28 de noviembre de 2018

Aceptado: 30 de noviembre de 2018

Correspondencia: fmendivelso@colsanitas.com

a relation to the estimator of the effect in the study. Our purpose was to present the clearest possible form, definition, uses and correct interpretation of *p-value* and *confidence intervals* 95%

**Keywords:** Confidence intervals; Hypothesis-Testing; Biostatistics; Data Interpretation, Statistical; Clinical Protocols

## EL p-valor

También denominado como “*valor crítico de la prueba estadística o nivel de significación alfa ( $\alpha$ )*” es ampliamente usado en la investigación clínica para el análisis de resultados (1). El *p-valor* es una cantidad que nos indica qué tan insólitos, extraños, raros o inusuales son los resultados en la muestra del estudio, considerando que la hipótesis nula ( $H_0$ ) establecida a priori sea verdadera. Dicho de otra forma, es la probabilidad (entre 0 y 1) de rechazar la  $H_0$  con los resultados obtenidos. Dado que rechazar una  $H_0$  cuando es verdadera es un error metodológico y estadístico grave, resulta razonable establecer a priori, que la probabilidad de cometerlo sea pequeña.

Si un investigador no puede rechazar la  $H_0$  cuando esta es verdadera o válida, significa que su investigación no ha eliminado una explicación alternativa importante para sus resultados. Rechazar la  $H_0$  no le dice nada al investigador sobre el mecanismo por el cual un factor causa o se relaciona con un desenlace particular o tamaño del efecto. No le dice si el estudio está bien diseñado o bien controlado ni tampoco si los resultados han sido seleccionados de forma sesgada; simplemente le ayuda a entender que tan raros son sus resultados y como lo raro de sus resultados se debieran solo al azar (2).

En la construcción del protocolo de investigación o diseño del estudio se propone un valor pequeño de  $\alpha$  para hacer que la probabilidad de rechazo de la  $H_0$  sea “razonable y clínicamente” pequeña. Los valores de  $\alpha$  más usados en investigación clínica son: 0,001 - 0,005 - 0,01 - 0,05 y 0,10. Sin embargo, no existe una regla determinada para seleccionar dicho *p-valor* (3). Lo más importante es entender que entre más pequeño sea  $\alpha$  o el valor crítico del estadístico de prueba, más exigente

es el diseño del estudio, la estimación del tamaño de la muestra, la técnica de muestreo, selección de la prueba estadística, control de sesgos y el análisis e interpretación de resultados (4).

## INTERPRETACIÓN INCORRECTA DEL p-valor

Es frecuente escuchar razonamientos erróneos en la presentación de resultados de investigación durante eventos académicos e incluso, en artículos científicos como el que mencionamos a continuación:

*“Con un  $\alpha$  de 0,05, nuestro estadístico de prueba para la diferencia de promedios obtuvo un valor de  $p = 0,07$ ; por lo tanto, no podemos “rechazar” la  $H_0$  y debemos “aceptar” la hipótesis alterna ( $H_1$ )”.*

En inferencia estadística (generar conclusiones a partir de una muestra “ $n$ ” a la población de referencia “ $N$ ”), es fundamental recordar que las pruebas de hipótesis ( $H_0$ ) requieren para su adecuada interpretación, el cálculo del tamaño de la muestra mediante un método probabilístico y que el estadístico de prueba se interpreta siempre; si y solo si, con relación a la  $H_0$  que es objetivo del estudio: se acepta o se rechaza la  $H_0$ . Ninguna mención se debe hacer sobre la  $H_1$  (5).

Otro error frecuente expuesto por muchos investigadores es mencionar que el nivel  $\alpha$  del 5% o 0,05, es el complemento de un Intervalo de Confianza (IC) del 95%. Como mencionamos antes, el investigador selecciona el nivel  $\alpha$  a priori para el contrástate de hipótesis con el estadístico de prueba más adecuado según la naturaleza de distribución de las variables de interés y puede fijar uno de varios valores, usualmente pequeños, y en nada corresponden al complemento de un IC95%. No

resulta ético, que en muchos estudios se trabaje a priori con un  $\alpha = 0,05$  y si los investigadores no encontraron resultados “estadísticamente significativos” incrementen el *p-valor* para “ajustar” sus resultados y presentar hallazgos positivos del estudio (6).

Las interpretaciones erróneas del *p-valor* han generado cuestionamientos serios sobre la relativa falsedad en las conclusiones de muchas investigaciones en medicina; incluso, han cuestionado la necesidad de seguir reportando el *p-valor* o de hacer más estricta su interpretación tratando de promover solo análisis con  $p < 0,005$ . Este mal entendido basado en la falsa seguridad de tener resultados con  $p < 0,05$  ha llevado a los investigadores y usuarios de literatura científica a tener una confianza excesiva en el *p-valor* (7).

### INTERPRETACIÓN CORRECTA DEL p-valor

Si el *p-valor* es muy pequeño (comparado con el  $\alpha$  establecido a priori), significa que los números raramente (¡Ojo... *pero no nunca!*) se producirán solo por casualidad o azar. De esta forma, cuando el *p-valor* es pequeño, los investigadores comienzan a pensar que la  $H_0$  parece improbable y dan un salto para concluir que “es muy poco probable que sus datos se deban al azar”.

*En general el p-valor no se debe usar para sacar conclusiones, sino para identificar posibilidades, como una prueba de esperanza matemática. Rechazar la  $H_0$  únicamente aporta evidencia indirecta de una hipótesis experimental. No dice nada acerca de si su conclusión científica es correcta o no. Esta máxima busca la cultura de no informar *p-valor* aislados, sino que en el reporte de los artículos científicos se incluya adicionalmente el tamaño del efecto y métricas de incertidumbre (Intervalos de Confianza) para mejorar no solo la reproducibilidad de la investigación científica, sino la adecuada interpretación de los hallazgos y conclusiones por la comunidad médica:*

*“Informar sobre los efectos que realmente no existen socava la credibilidad de la ciencia” (8)*

### INTERVALOS DE CONFIANZA

Un intervalo de confianza (IC) describe la *variabilidad* entre la medida de efecto (RA, RR, OR, HR, NNT, etc.) obtenida en un estudio y la medida real de la población (el valor real). El IC corresponde a un *rango de valores*, cuya distribución es normal y en el cual se encuentra, con alta probabilidad, el valor real de una determinada variable. Esta “*alta probabilidad*” se ha establecido por consenso en 95%. Así que, un IC95% indica que dentro del rango dado (un número superior e inferior calculado a partir de una muestra) se encuentra el valor real de un parámetro con 95% de certeza (9).

Debido a que la medida de efecto real de la población es desconocida, este rango describe los valores posibles que podría tener la medida de efecto si se tomaron muestras múltiples de la misma población y se calcula nuevamente un IC95% para cada muestra; luego, esperaríamos que la medida del efecto en la población (N) se encontrara dentro del 95% de estos IC (5).

Un IC es sensible a la variabilidad y al tamaño de la muestra. Es necesario advertir que existe una relación inversa entre la amplitud del IC y el *tamaño muestral* ( $n$ ): mientras más pequeño es la  $n$ , más amplio es el IC. A mayor  $n$  más estrecho el IC y más certeza tendremos que el resultado del estudio se acerca al valor real.

Cuando comparamos las medidas de efecto entre dos o más grupos de análisis o intervención (experimentos), el IC muestra la verdadera magnitud de la diferencia entre grupos. Esto resulta extremadamente útil para comprender tanto la importancia estadística como clínica de un tratamiento o intervención. Un IC95% da más información que un *p-valor* dado que le permite al investigador, comparar la magnitud de una diferencia y permite entender si la valoración de significancia o no significancia estadística es consecuencia de la estimación del tamaño de la muestra (10).

### CONCLUSIÓN

Al redactar el informe final de un manuscrito de investigación clínica, se deben presentar las estimaciones de

los efectos con sus respectivos IC95% y no calificar o redactar en las diferentes secciones del artículo, que el *p*-valor fue “*significativo*” o “*no significativo*”, únicamente informarlos. El IC95% es una medida de precisión que

permite al clínico evaluar si una diferencia reportada es relevante para recomendarla a sus pacientes (relevancia clínica).

## REFERENCIAS

1. Chavalarias D, Wallach JD, Li AHT, Ioannidis JPA. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. JAMA. 2016 Mar;315(11):1141-8.
2. Daniel WW, Cross CL. Biostatistics: a foundation for analysis in the health sciences. Wiley; 2018.
3. Consonni D, Bertazzi PA. Health significance and statistical uncertainty. The value of P-value. Med Lav. 2017 Oct;108(5):327-31.
4. Mendivelso F, Rodríguez M. Independence Chi-Square test applied to 2xN Tables. RevMedicaSanitas. 2018;21(2):92-5.
5. O'Brien SF, Yi QL. How do I interpret a confidence interval? Transfusion. 2016 Jul;56(7):1680-3.
6. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005 Aug;2(8):e124.
7. Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. JAMA. 2018 Apr;319(14):1429-30.
8. Hayden EC. Weak statistical standards implicated in scientific irreproducibility. Nat News. 2013;10:14131.
9. Dunst CJ, Hamby DW. Guide for calculating and interpreting effect sizes and confidence intervals in intellectual and developmental disability research studies. J Intellect Dev Disabil. 2012 Jun;37(2):89-99.
10. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007 Nov;82(4):591-605.