

Valor p , intervalo de confianza y significancia clínica

P-value, confidence interval and clinical significance

Señor Editor,

Hemos leído con gran interés la editorial “Interpretando al *valor p*: ¿Significación estadística o significación clínica?”¹ publicada en el No. 2 de 2024 de *Odontología Sanmarquina*, en la que se aborda un tema de enorme pertinencia e importancia en el ámbito de la bioestadística que amerita ser celosamente aplicado en la investigación biomédica. La autora explica con sumo cuidado el *valor p* y recomienda que la interpretación del resultado se haga a la luz del tamaño del efecto y de la muestra, aparte de otras exigencias asociadas con los modelos estadísticos empleados en cada caso particular. De ser así, afirma ella, se podría constatar que las diferencias estadísticamente significativas encontradas corresponden muy probablemente a verdaderas diferencias clínicamente significativas. Conocer esto por adelantado evitaría el error de asumir significancia clínica siempre que el *valor p* está por debajo del nivel de significancia estadística preestablecido.

Para ampliar el excelente aporte de la autora, los autores de este escrito detallaremos el proceso de significancia clínico-estadística mediante un ejemplo. Suponga que se quiere establecer si existe diferencia entre el índice de cariosos, perdidos y obturados (CPO) según el sexo (hombre/mujer) a fin de definir la necesidad de intervención en una población. El proceso estadístico para establecer si existe diferencia en el COP entre los sexos, en la población sería el siguiente:

Jorge Homero Wilches-Visbal¹, Midian Clara Castillo-Pedraza¹, Kelly Johana Obispo-Salazar¹

¹ Universidad del Magdalena, Facultad de Ciencias de la Salud, Santa Marta, Colombia.

Correspondencia:

Jorge Homero Wilches Visbal: jhwilchev@gmail.com
Universidad del Magdalena, Laboratorio de Biofísica,
Edificio de Innovación y Emprendimiento. Ciudadela
Universitaria, Calle 29H3 No 22 – 01, San Pedro
Alejandrino, Santa Marta, Colombia
ORCID: 0000-0003-3649-5079

Coautores:

Midian Clara Castillo-Pedraza:
mcastillo@unimagdalena.edu.pe
ORCID: 0000-0003-3170-3959
Kelly Johana Obispo-Salazar:
kobispos@unimagdalena.edu.pe
ORCID: 0000-0003-3618-1192

Conflictos de intereses: Los autores declaran no tener ningún conflicto de interés.

Fuente de financiamiento: Autofinanciado.

Recibido: 16/08/2024
Aceptado: 28/01/2025
Publicado: 31/03/2025

1. *Muestra representativa*: como es imposible acceder a la totalidad de la población, se selecciona una porción de esta (muestra) que idealmente sea tan grande (muchas personas) y heterogénea (similar en proporción de rasgos sociodemográficos) como la población de la que proceden. Así, se busca contar con la mayor cantidad de participantes posibles, seleccionados de forma aleatoria. Para esto, se podría optar por ubicarse en el centro de la ciudad de interés en el estudio y pedir la participación voluntaria de las personas que transitan. Entre más representativa la muestra, más *probable* es que el valor estimado de la variable objetivo se aproxime a su verdadero valor en la población. Implícitamente, implica pasar de la esfera de la certeza (resultado poblacional, que “solo lo sabe Dios”) a la de la probabilidad (resultado muestral, “al que puede acceder el humano”).
2. *Prueba de hipótesis*: se supone de entrada una hipótesis que es cierta, la hipótesis nula (H_0). En dicho caso: “no existen diferencias del índice CPO entre hombres y mujeres en la población”. Posteriormente se aplica una prueba estadística de significancia indicada para el caso, en el ejemplo, una *t* de *Student* (o Mann-Whitney), a fin de corroborar o rechazar (con cierta probabilidad de error o con seguridad razonable) la H_0 en favor de la hipótesis alternativa (H_a), que es lo que normalmente se pretende.
3. *Valor p*: se compara el *valor p* dado por la prueba estadística con un límite de criticidad preestablecido denominado nivel de significancia, el cual suele ser 0,05 o 5% ². En el hipotético caso de que el *valor p* del ejemplo diera 0,01, esto significaría que “la probabilidad de haberse equivocado al rechazar la H_0 cuando esta se supuso verdadera es 1%”. En otras palabras, como es tan poco probable que H_0 sea cierta (no falsa, porque se está en el terreno de lo probable), dado que $p = 0,01 < 0,05$, entonces se le rechaza en favor de H_a ³. Dado que el *valor p* encontrado es más extremo que el límite de criticidad es bastante probable que de verdad haya diferencia del CPO entre los sexos, en la muestra. Además, si esta última es verdaderamente representativa, entonces la diferencia es muy probable que también exista en la población (se recomienda inspeccionar el poder estadístico de la prueba). Por esta razón, al *valor p* suele definírsele como la probabilidad de que la diferencia observada obedezca al azar. Gráficamente se podría entender como el área de superposición (en porcentaje) entre la distribución estadísticas de la variable objetivo de los grupos de comparación (en este caso, el porcentaje de superposición de la distribución de COP entre los sexos) ³.
4. *Significancia estadística vs significancia clínica*: se supone que, en el ejemplo, el *valor p* fue 0,01, por tanto, se concluye que hay diferencias estadísticamente significativas entre el índice CPO de hombres y mujeres. Ahora bien, ¿esto tiene verdaderamente una implicación desde el punto de vista clínico? en otras palabras, ¿la diferencia entre el índice CPO de

mujeres es tan inferior/superior al de hombres que amerita una intervención diferencial (en términos odontológicos o de salud pública) en la población objetivo? Es aquí donde entra el *tamaño del efecto*: demostrada la diferencia, determine qué tan grande es la magnitud de la misma y si esta es lo suficiente como para que tenga relevancia clínica ⁴.

Por otro lado, ¿qué habría sucedido si el *valor p* hubiera sido mayor al nivel de significancia ($p > 0,05$)? La respuesta es que se habría concluido que “no hay diferencias estadísticamente significativas entre el índice CPO de mujeres y hombre en la población de la cual provienen (o población objetivo). Aunque lo esperable es que cuando se verifica la existencia de diferencias estadísticamente significativas, se confirme también la diferencia clínicamente significativa, también es posible que esta se dé cuando la primera no ³.

Molina³ sugiere utilizar el intervalo de confianza al 95 % (IC95) para descubrir diferencias clínicamente significativas en lugar del *valor p*. Para ejemplificar, suponga que el umbral de significancia clínica (previamente establecido mediante estudios, experticia del profesional, guías o grupo de expertos, etc) es aquel en el que la diferencia entre los índices CPO de hombres y mujeres es de 10 (puede normalizarse por el desvío estándar para fines de comparación con el tamaño del efecto). Al respecto, pueden recrearse dos escenarios hipotéticos: i) IC95 con límite inferior de 0 y límite superior de 11: *no* hay diferencia estadísticamente significativa porque el IC contiene al 0 pero parece haber diferencia clínica porque el límite superior es mayor que 10, en dicho caso se recomienda ampliar la muestra; ii) IC95 con límite inferior de 2 y superior de 5: *hay* significancia estadística (el IC no contiene al 0) pero *no* clínica porque el límite superior del IC es inferior a 10. Esto se traduce en que, si bien se observa una diferencia estadísticamente significativa en el índice CPO entre hombres y mujeres, la magnitud de dicha diferencia podría no ser clínicamente relevante ni justificar, por sí sola, intervenciones diferenciadas entre los grupos. Esta interpretación es válida siempre que la muestra sea representativa de la población de interés, considerando además la evaluación del poder estadístico de la prueba aplicada.

En conclusión, el *valor p* en sí mismo no basta para afirmar o negar los efectos reales de una intervención o verificar diferencias entre dos grupos respecto a una variable de estudio, es necesario además establecer a priori el umbral de significancia clínica (según la experticia del profesional o basado en investigaciones previas) así como constatar si se alcanzó y, para ello, han de tenerse en cuenta parámetros como el tamaño del efecto y/o el intervalo de confianza.

Referencias

1. Rodríguez-Vargas, M. Interpretando al valor p: ¿Significación estadística o significación clínica? *Odontol Sanmarquina*. 2024; 27(2): e27997. DOI: <https://doi.org/10.15381/os.v27i2.27997>

2. Martínez-Ezquerro J.D., Riojas-Garza A. y Rendón-Macías M.E. Significancia clínica sobre significancia estadística. Cómo interpretar los intervalos de confianza a 95 %. Rev Alerg México. 2017; 64(4): 477–86. DOI: <https://doi.org/10.29262/ram.v64i4.334>
3. Molina Arias, M. ¿Qué significa realmente el valor de p? Pediatr Aten Primaria. 2017;19(76): 377–81. Disponible en: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322017000500014
4. Kain, Z. y MacLaren, J. Valor de p inferior a 0.05: ¿qué significa en realidad? Pediatr (Ed. Española). 2007; 63(3): 118–20. Disponible en: <https://www.el-sevier.es/es-revista-pediatrics-10-articulo-valor-p-inferior-005-que-13112660>