

Los lectores de Medicina Clínica conocen la importancia de definir el denominador de una proporción para estimar una probabilidad.

El valor p (o valor de p , o p -valor, o simplemente p) guarda cierta analogía con las probabilidades diagnósticas, ya que se define como la probabilidad de obtener un resultado tan significativo o más que el observado. No obstante, a un investigador o a un clínico le puede resultar más interesante conocer el valor positivo de una prueba.

En investigación, la replicabilidad y transparencia de un experimento es fundamental, ya que nos acerca al rigor científico: cualquier investigador independiente debería poder replicar los resultados.

El uso del valor de p y la significación estadística han estado en entredicho desde principios de la década de los 80 en el siglo pasado hasta nuestros días.

Mucho se ha discutido al respecto en el ámbito de la estadística y sus aplicaciones, en particular a la Epidemiología y la Salud Pública.

Sin embargo, su significado debería ser claro en términos intuitivos a pesar de que se basa en conceptos teóricos del terreno de la Estadística-Matemática.

En esta edición, Salus ha seleccionado para el Tópico de Actualidad a Luis Pérez-Ybarra del Departamento de Ciencias Básicas de la Escuela de Bioanálisis, Sede Aragua y a William Quintana-Rivero del Departamento de Matemática, Estadística y Técnicas Cuantitativas, Escuela de Administración Comercial y Contaduría Pública del Campus La Morita, Maracay, Venezuela, para tratar este interesante tema de actualidad

Comité Editorial Salus



<https://doi.org/10.54139/salus.v25i2.56>

La controversia acerca del uso del p -valor en el mundo científico.

En la actualidad existe una controversia en el mundo científico en relación al uso del p -valor (p -value, en inglés) como herramienta de decisión para llevar a cabo pruebas de hipótesis estadísticas, en el presente ensayo se mostrará de forma muy sucinta en qué consiste esta controversia.

Cuando un investigador decide trabajar en algún proyecto, por lo general, parte de una hipótesis de trabajo, y desea verificar si dicha hipótesis está apoyada o no por la evidencia empírica. Si la hipótesis está apoyada por la evidencia empírica, se considera a la hipótesis como cierta, y en caso contrario como falsa, para hacer eso, lo usual es extraer una muestra de la población objeto de estudio; la forma de extraerla dependerá del tipo de investigación, aunque las metodologías más utilizadas son realizar un muestreo aleatorio o intencional, conducir un experimento, o tomar de datos previamente archivados. En cualquier caso, es fundamental que la muestra sea representativa de la población objeto de estudio, ya que esta propiedad garantiza la confiabilidad de los resultados obtenidos.

Una vez seleccionada la muestra, sobre ésta se procede a medir variables de interés para la investigación, para tal fin se utilizan instrumentos de recolección de datos como la encuesta, la aplicación de una prueba estandarizada o la medición de ciertas variables de interés mediante

instrumentos o pruebas de laboratorio. Por supuesto, se supone que estas variables de interés están relacionadas y dan respuesta a los objetivos de la investigación, y en última instancia, permiten verificar si la hipótesis de trabajo se mantiene o no. Frecuentemente ocurre que dichas variables de interés son de naturaleza cuantitativa, y el problema se reduce al análisis estadístico de estas variables. Así las cosas, la hipótesis de trabajo se operacionaliza en una serie de hipótesis estadísticas formuladas sobre variables cuantitativas medidas en los datos recolectados en la muestra seleccionada.

La estadística, en virtud de esta necesidad de decidir si una hipótesis se puede considerar como verdadera o falsa, ha desarrollado un mecanismo de toma de decisiones conocido como pruebas de hipótesis. En ese sentido, las pruebas de hipótesis son una regla de decisión que permite decidir entre una hipótesis nula (H_0) y una hipótesis alternativa (H_1), donde la H_0 y la H_1 constituyen un sistema de hipótesis de las cuales el investigador seleccionará solo una. En forma general, se procede de la siguiente manera: se establece una H_0 , la cual corresponde a algún valor (o rango de valores) o característica de un parámetro desconocido de la población, y una H_1 , que es complementaria a la hipótesis nula, de tal forma que ambas dividan el espacio muestral de algún estadístico de prueba asociado en dos zonas distintas: una zona de aceptación de H_0 y una zona de rechazo de H_0 , y partiendo del valor de una transformación de los datos en el estadístico de prueba.

Por ejemplo, en el estadístico *t* de Student, se toma la decisión de aceptar la H_0 o la H_1 . Es necesario aclarar que el estadístico de prueba presenta una distribución de probabilidad conocida y que los valores obtenidos en la muestra siguen tal distribución.

La decisión de aceptar la H_0 o la H_1 se toma a partir de un valor numérico conocido como *p*-valor, y éste corresponde a la probabilidad de que se pueda obtener un valor tan o más extremo que el valor del estadístico de prueba obtenido en la muestra seleccionada.

Un *p*-valor muy pequeño indica que, suponiendo que la hipótesis nula sea cierta, es poco probable su ocurrencia en la muestra analizada (o una muestra cualquiera), lo que ofrece evidencia de que probablemente la H_0 sea falsa, y se deba decidir en favor de la H_1 .

El valor a partir del cual se puede considerar lo suficientemente pequeño al *p*-valor se denomina nivel de significación, y aunque no hay una razón práctica o teórica para su selección, usualmente se fija en 0,05 o 0,01. La regla de decisión usual es considerar que si el $p\text{-valor} \leq 0,05$ (o $p\text{-valor} \leq 0,01$) hay significación estadística, y se decide que la H_0 es falsa y se toma en su lugar a la H_1 como cierta.

Lamentablemente, si bien el procedimiento operativo general es el descrito, las cosas no son tan simples como parecen. El uso indiscriminado, mecánico, descuidado y poco fundamentado de este criterio ha llevado muchas veces a la toma de decisiones e interpretaciones erróneas de los resultados, a tal punto que, en la actualidad, está en discusión si es viable o no utilizar al *p*-valor como herramienta para la toma de decisiones estadísticas.

La ASA (American Statistical Association por sus siglas en inglés) (1) en el año 2016 publicó una serie de observaciones sobre el significado y el uso del *p*-valor en la toma de decisiones estadísticas o pruebas de hipótesis. Las mismas corresponden básicamente a aclaratorias sobre los errores más comunes cometidos al interpretar el concepto de *p*-valor. La ASA señala que el *p*-valor solo indica que tan incompatibles son los datos con el modelo matemático asumido como cierto (H_0), por lo cual el *p*-valor por sí mismo no indica la probabilidad de que la hipótesis H_0 sea cierta o falsa o que los resultados sean debidos al azar. Simplemente es un valor de referencia para tomar una decisión entre H_0 y H_1 , con todos los riesgos que esto conlleva, pero nada dice en realidad sobre la certeza o la naturaleza de H_0 o H_1 .

Asimismo, recomienda que las decisiones tomadas no se basen únicamente en el *p*-valor, sino que vengan acompañadas de una estimación del efecto mediante la

construcción e interpretación de los intervalos de confianza adecuados, y de un análisis cuidadoso y concienzudo de las condiciones metodológicas asociadas a la conducción de la investigación, la toma de muestra, las hipótesis seleccionadas y los *p*-valores calculados y seleccionados para su publicación.

Se subraya que el *p*-valor no constituye una estimación del tamaño del efecto: el tamaño de un efecto puede ser grande o pequeño independientemente del tamaño del *p*-valor, ya que este último es función del tamaño del efecto, de la varianza de los datos y del tamaño de la muestra. Además, se advierte que la significación estadística no necesariamente va de la mano con la utilidad práctica.

En el caso de las ciencias biomédicas se ha vertido mucha tinta distinguiendo entre la significación clínica y la significación estadística (2). Puede darse el caso de que una variable presente significación estadística pero no sea relevante desde el punto de vista clínico, y a su vez, puede ocurrir que un efecto medido no sea estadísticamente significativo, pero pudiera ser clínicamente importante.

Se considera que la confiabilidad del *p*-valor va de la mano del apego del investigador a la aplicación de una metodología correcta y escrupulosa y si no se garantiza el cumplimiento de los supuestos o hipótesis previas para la aplicación de las pruebas de hipótesis: independencia de las observaciones, ajuste a la distribución normal, representatividad de la muestra, reproducibilidad de las observaciones, y control local de posibles variables concomitantes (covariables modificadoras de la respuesta, ruido estadístico en general, imposibilidad de contar con observaciones medidas bajo condiciones idénticas, por ejemplo).

Entonces el *p*-valor no sería garantía para una correcta toma de decisiones, lo cual a la postre, es para lo que fue diseñado, no para seleccionar de forma infalible a la hipótesis correcta, y por tanto, como toda decisión realizada bajo incertidumbre lleva asociado un riesgo de error, el cual, si se garantizan las condiciones necesarias, puede reducirse y controlarse. No es casual que muchas de las objeciones al uso del *p*-valor sean de naturaleza metodológica e interpretativa, sobre todo asociadas a la reproducibilidad y replicabilidad de los resultados y a las conclusiones derivadas de una investigación (3), hasta tal punto que algunas revistas científicas no recomiendan, o prohíben abiertamente el uso del *p*-valor en sus publicaciones (4, 5).

La discusión sigue abierta y se dista mucho de llegar a un consenso. Por una parte, muchos investigadores consideran que el *p*-valor puede utilizarse, pero con cautela y con un basamento metodológico sólido, o tomando como nivel

de significación a 0,005 en vez del usual 0,05 (6). Otros investigadores proponen enfoques alternativos para probar hipótesis basados en métodos de estimación bayesianos (7) y, por supuesto, hay investigadores que consideran sin ambages, que el uso del p-valor debe ser eliminado de la investigación científica y de la toma de decisiones (8).

En cualquier caso, históricamente el p-valor ha mostrado ser una herramienta útil y muy potente cuando se aplica con seriedad y será una fuente de errores y malas interpretaciones cuando se usa a la ligera. Solo el tiempo dirá como terminará o evolucionará esta controversia, que nos atañe a todos los que de una u otra manera estamos ligados a la investigación científica.

REFERENCIAS BIBLIOGRÁFICAS

1. Wasserstein RL, Lazar NA. The ASA statement on p-Values: context, process, and purpose. *Am Stat.* 2016; 70(2):129-133. DOI: 10.1080/00031305.2016.1154108
2. Molina Arias M. ¿Qué significa realmente el valor de p? *Rev Pediatr Aten Primaria.* 2017; 19(76):377-381.
3. Badenes-Ribera L, Frías Navarro ML. Falacias sobre el valor p compartidas por profesores y estudiantes universitarios. *Universitas Psychologica.* 2017; 16(3):1-10. DOI: 10.11144/Javeriana.upsy16-3.fvcp
4. Monterrey Gutiérrez P. $p < 0,05$, ¿Criterio mágico para resolver cualquier problema o leyenda urbana? *Universitas Scientiarum.* 2012; 17(2):203-215. DOI: 10.11144/javeriana.SC17-2.pamc
5. Fricker RD Jr, Burke K, Han X, Woodall WH. Assessing the statistical analyses used in Basic and Applied Social Psychology after their p-value ban. *Am Stat.* 2019; 73(Suppl. 1):374-384. DOI: 10.1080/00031305.2018.1537892
6. Niz-Ramos J. Las falacias de la p y significación estadística. *Ginecol Obstet Mex.* 2020; 88(8):536-541. DOI: 10.24245/gom.v88i8.4534
7. Pérez González JD, Frías-Navarro ML. Retract $p < 0.005$ and propose using JASP, instead. *F1000Research.* 2018; 6:2122. DOI: 10.12688/f1000research.13389.2
8. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat.* 2019; 73(Suppl. 1):235-245. DOI: 10.1080/00031305.2018.1527253

Luis Pérez-Ybarra

<https://orcid.org/0000-0003-0743-7953>

Email: Impy2005@gmail.com

Departamento de Ciencias Básicas, Escuela de Bioanálisis sede
Aragua,
Facultad de Ciencias de la Salud, Universidad de Carabobo,
Venezuela

William Quintana-Rivero

<https://orcid.org/0000-0003-4369-6557>

Email: wquintanar@gmail.com

Departamento de Matemática, Estadística y Técnicas
Cuantitativas. Escuela de Administración Comercial y Contaduría
Públicas. Sede Aragua. Facultad de Ciencias Económicas y
Sociales. Universidad de Carabobo. Maracay. Venezuela

Salus