

Facultad de Medicina



## Magnitud del efecto para pruebas de normalidad en investigación en salud

Effect size for normality test in health research



Sr. Editor:

Las medidas de magnitud del efecto (ME) suelen ser asociadas con procedimientos empíricos como comparación entre grupos, correlaciones, entre otros<sup>1</sup> como una alternativa a la interpretación aislada de la *prueba de significancia de la hipótesis nula* (NHST, por sus siglas en inglés). Sin embargo, si bien las ME se orientan a la valoración del *tamaño de la diferencia* entre grupos<sup>2</sup> o la *fuerza de asociación* entre dos o más variables<sup>3</sup>, también podría utilizarse en procedimientos vinculados al análisis de normalidad (p.e., con la prueba de Kolmogorov-Smirnov; K-S), cuya interpretación depende de la NHST.

Esto es relevante porque frecuentemente la decisión sobre el cumplimiento o no de la normalidad, con base en la NHST, tiene dos aspectos a considerar. El primero, es que el reporte dicotómico del análisis de la normalidad (*cumple / no cumple*) no informa qué tan cercana o lejana se halla la distribución empírica de la distribución normal, ya que la hipótesis nula ( $H_0$ : *la variable X se aproxima a una distribución normal*) no será retenida si el *p*-valor asociado al estadístico de prueba es menor que el *error tipo I* ( $\alpha$ ), sin considerar otros criterios complementarios.

El segundo aspecto se vincula con la anterior, ya que la exigencia de distribución normal para puntuaciones que reflejan constructos psicológicos usados en investigación en salud (p.e., ansiedad, depresión) es poco realista<sup>4,5</sup>, y si no se retiene la  $H_0$  podrían utilizarse procedimientos analíticos que, probablemente, provean resultados que representen una pérdida de información al pasar de un nivel de medición a otro (p.e., la *U* de Mann-Whitney convierte las puntuaciones directas a medidas ordinales denominadas *rangos*) o que puedan conducir a una modificación implícita de los objetivos de la investigación

debido a que algunos estadísticos aparentemente equivalentes proveen información distinta (p.e., el coeficiente de correlación de Pearson evalúa la relación lineal entre las variables, mientras que el de Spearman analiza si la relación es monotónica). Ante esa circunstancia, existen estudios que se establecen criterios para utilizar determinados estadísticos paramétricos aun cuando no se cumple el supuesto de normalidad<sup>6</sup>.

Por tal motivo, además del cumplimiento o no de la normalidad es necesario informar qué tanto se alejan los datos de dicha distribución. En este punto, es útil el concepto de ME dado que se desea conocer la *magnitud del alejamiento*. En tal sentido, cuando se decide el uso de la prueba K-S-L (K-S con corrección Lilliefors) (una de las más usadas para evaluar las características distribucionales de los datos a pesar sus limitaciones<sup>7</sup>) la medida de ME es la *diferencia más extrema absoluta* (*D*) asociada al estadístico *Z* de dicha prueba. El cálculo de *D* es automático en paquetes estadístico comerciales (p.e., SPSS<sup>(R)</sup>, donde aparece como *Most Extreme Differences - Absolute*), pero resulta esclarecedor conocer la expresión matemática para derivar la propuesta de valoración:  $D = Z / \sqrt{n}$ , siendo *n* el tamaño muestral.

Esta expresión es similar a la utilizada para calcular la ME cuando se emplea la *U* de Mann-Whitney ( $r = Z / \sqrt{n}$ ) para analizar desde un enfoque no paramétrico si las distribuciones de dos grupos son estadísticamente similares<sup>1</sup>. Por tal motivo, y haciendo una extrapolación de los puntos de corte propuestos para *r*<sup>8</sup> en vista que ambas son pruebas no paramétricas y comparan distribuciones, *D* podría valorarse como una *medida de alejamiento de la distribución normal* según los siguientes puntos de corte: menor que .10, *alejamiento*

La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

to insignificante, entre .10 y .30, *pequeño*; entre .30 y .50, *moderado*; y  $> .50$ , *grande*. Además, cuando  $n$  es pequeño, la interpretación de la  $D$  evitaría la sobreestimación del acercamiento a la normalidad, ya que como se conoce en estos casos es más probable retener la  $H_0$ .


Por ejemplo, en un estudio sobre *indecisión para elegir la carrera profesional y autoeficacia percibida*<sup>9</sup>, previamente se analizó si las puntuaciones utilizadas se aproximan a la normalidad o no. En dicha investigación, las dimensiones de *información inconsistente* ( $Z = 2.262$ ;  $p < .001$ ;  $n = 82$ ) y *escases de información* ( $Z = 4.299$ ;  $p < .001$ ;  $n = 108$ ) no se aproximan a la normalidad desde la NHST. Asimismo, las  $D$  fueron de .250 y .414, respectivamente. Es decir, mientras que el primero podría calificarse como un alejamiento *cercano a moderado*, y el segundo como *moderado*. A su vez, en ese mismo estudio la *autoeficacia percibida* se aproxima a la normalidad ( $Z = 1.285$ ;  $p = .073$ ;  $n = 28$ ), aunque su alejamiento evaluado con la  $D$  es *cercano a moderado* ( $D = .243$ ). Esta situación ilustraría que, independiente del resultado derivado de la NHST, es necesario reportar e interpretar  $D$ .

Entonces, con esta información el investigador evaluaría si el grado de alejamiento de la normalidad que presentan sus datos (p.e., *insignificante* o *pequeño*) es *tolerable*, según el diseño de investigación elegido o el constructo estudiado, para decidir por análisis paramétricos o no paramétricos. Asimismo, cabe precisar que a pesar que K-S-L no es la única prueba de normalidad disponible<sup>7</sup>, tiene un mejor desempeño con muestra pequeñas y moderadas en distribuciones simétricas, pero no mejor que la prueba de Shapiro-Wilk que funciona adecuadamente con muestras grandes ( $n > 500$ )<sup>10</sup>. Los estudios citados previamente<sup>7,10</sup> son importantes para decidir qué prueba usar ya que se basan en la capacidad de cada una para rechazar la  $H_0$ , e incluso uno de ellos realiza la simulación considerando la presencia de distribución normal para todas las muestras generadas<sup>10</sup>. Sin embargo, estos no hacen referencia al *grado de alejamiento de la normalidad* que posee la distribución de los datos, que representa el objetivo del presente manuscrito.

Finalmente, al ser una primera aproximación para

la valoración cuantitativa de  $D$  como una medida de ME, es necesario realizar estudios que permitan establecer puntos de corte con base empírica como ya fue realizado con otras ME con el objetivo de valorar la  $D$  en el contexto de la investigación realizada<sup>11</sup>.

## CONFLICTO DE INTERESES

Ninguno. 

## REFERENCIAS

1. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen*. 2012; 141(1): 2-18. doi: 10.1037/a0024338
2. Dominguez-Lara S. Magnitud del efecto en comparaciones entre dos o más grupos. *Revista de Calidad Asistencial* 2017; 32(2), 121 -2. doi: 10.1016/j.cali.2016.04.002
3. Dominguez-Lara S. Magnitud del efecto en análisis de regresión. *Interacciones* 2017; (1), 3-5. doi: 10.24016/2017.v3n1.46
4. Micceri T. The unicorn, the normal curve, and other improbable creatures. *Psychol Bull* 1989; 105: 156 – 66. doi: 10.1037/0033-2909.105.1.156
5. Ho AD, Yu CY. Descriptive Statistics for Modern Test Score Distributions Skewness, Kurtosis, Discreteness, and Ceiling Effects. *Educ Psychol Meas* 2015; 75(3): 365-88. doi: 10.1177/0013164414548576
6. De Winter JCF, Gosling SD, Potter J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol Methods* 2016; 21(3): 273 – 90. doi: 10.1037/met0000079
7. Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab* 2012; 10(2): 486-9.
8. Coolican H. Research methods and statistics in psychology. London: Hodder; 2009.
9. Crisan C, Turda S. The connection between the level of career indecision and the perceived self-efficacy on the career decision-making among teenagers. *Procedia* 2015; 209: 154 – 60. doi: 10.1016/j.sbspro.2015.11.271
10. Pedrosa I, Juarros-Basterretxea J, Robles-Fernández A, Basteiro J, García-Cueto E. Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar? *Univ Psychol* 2015; 14(1): 245-54. doi: 10.11144/Javeriana.upsy13-5.pbad
11. Bosco FA, Aguinis H, Singh K, Field JG, Pierce CA. Correlational effect size benchmarks. *J Appl Psychol* 2015; 100(2), 431 – 49. doi: 10.1037/a0038047

Sergio Dominguez-Lara<sup>a</sup>

<sup>a</sup>Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Lima, Perú.

\*Correspondencia: Sergio Alexis Dominguez Lara, Av. Tomás Marsano 242, Surquillo – Lima 34 (Perú).

C. electrónico: sdominguezl@usmp.pe, sdominguezmpcs@gmail.com

<http://dx.doi.org/10.22201/facmed.20075057e.2018.27.1776>  
2007-5057/© 2017 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).