

ERRORES DE INTERPRETACIÓN DE LOS VALORES p ENTRE PSICÓLOGOS PROFESIONALES ESPAÑOLES: UN ESTUDIO EXPLORATORIO

Laura Badenes-Ribera,

Profesor Asociado del Departamento de Metodología de las Ciencias del Comportamiento.

Facultad de Psicología.

laura.badenes@uv.es

Dolores Frias-Navarro,

Amparo Bonilla-Campos

Universitat de València

<https://doi.org/10.17060/ijodaep.2017.n1.v2.870>

Fecha de Recepción: 23 Marzo 2017

Fecha de Admisión: 1 Abril 2017

RESUMEN:

Este artículo analiza los errores de interpretación del valor p cometidos por los psicólogos profesionales españoles dado que la falta de conocimiento y comprensión del valor de p puede afectar las decisiones de los profesionales y poner en peligro la calidad de las intervenciones psicológicas. Se realizó una encuesta a 77 psicólogos profesionales españoles (68,8% mujeres, con una edad media de 41.44 años, DT = 9,42). Nuestros hallazgos son consistentes con previas investigaciones y sugieren que muchos participantes no saben cómo interpretar correctamente los valores de p . Más de un tercio de ellos creyó que el valor p indica la importancia clínica o práctica de los hallazgos. Estos resultados destacan la importancia de la formación en conceptos estadísticos, metodología de diseño de investigación y pruebas de inferencia estadística con el fin de mejorar la práctica profesional entre los psicólogos profesionales.

Palabras clave: valores de p , concepciones erróneas, tamaño del efecto, falacias.

ABSTRACT:

This paper analyzes the misinterpretations of the p -value made by Spanish practitioner psychologists given that lack of knowledge and understanding of p value might affect the professionals' decisions and jeopardize the quality of psychological interventions. We carried out a survey on 77 practitioner psychologists from Spain (68.8% women, mean age of 41.44 years, SD = 9.42). Our findings are consistent with previous research and suggest that many participants did not know how to correctly interpret p values. More than a third of them believed that p -value indicate the clinical or practical significance of the findings. These results highlight the importance of the training in sta-

tistical concepts, research design methodology, and statistical inference tests in order to improve professional practice among practitioner psychologists.

Keywords: *p*-value, misconceptions, effect size, fallacies.

ANTECEDENTES

La Práctica Basada en la Evidencia (PBE) se define como “*la integración de la mejor evidencia disponible con la experiencia clínica en el contexto de las características, cultura y preferencias del paciente*” (American Psychological Association (APA), Presidencial Grupo de Trabajo sobre la Práctica Basada en la Evidencia, 2006, p. 273). Por definición, la PBE se basa en la utilización de la investigación científica en la toma de decisiones en un esfuerzo por producir los mejores servicios posibles en la práctica clínica (APA, 2005; Babione, 2010; Daset y Cracco, 2013; Sánchez-Meca, Boruch, Petrosino, y Rosa-Alcázar, 2002; Vázquez y Nieto, 2003). Por tanto, el enfoque de la PBE requiere de los profesionales nuevas habilidades como la capacidad para evaluar y jerarquizar la calidad de las investigaciones psicológicas, para proporcionar el mejor servicio posible a los pacientes mediante la incorporación de la mejor evidencia en la experiencia o el juicio profesional, junto a las opiniones de los pacientes (Beyth-Marón, Fidler, y Cumming, 2008; Authors, 2003; Sackett, Straus, Richardson, Rosenberg, y Haynes, 2000).

Dentro de este proceso de evaluación crítica de la evidencia es crucial conocer y comprender el proceso de la prueba de significación de la hipótesis nula (*Null Hypothesis Significance Testing*, NHST) como herramienta para el análisis de datos, dado que este procedimiento goza de una considerable difusión en la investigación en Psicología, siendo utilizado en la mayor parte de los artículos publicados en revistas del área (Cumming y cols., 2007). En consecuencia, saber cómo interpretar los valores *p* de probabilidad es una competencia básica del profesional en Psicología y en cualquier disciplina en que se aplique la inferencia estadística.

El valor de *p* relacionado con los resultados de una prueba estadística es la probabilidad de obtener los datos observados o un valor más extremo si la hipótesis nula es verdadera (Kline, 2013). La definición es clara y precisa, sin embargo, los conceptos erróneos de los valores *p* siguen siendo numerosos y repetitivos (Autor, 2015, 2016, 2017; Falk y Greenbaum, 1995; Haller y Krauss, 2002; Kühberger, Fritz, Lerner, y Scherndl, 2015; Oakes, 1986).

Los errores de interpretación más comunes del valor *p* son la “falacia de la probabilidad inversa”, la “falacia de la replicación”, la “falacia del tamaño del efecto” y la “falacia de la significación clínica o práctica” (Carver, 1978; Cohen, 1994; Harrison, Thompson, y Vannest, 2009; Kline, 2013; Nickerson, 2000; Wasserstein y Lazar, 2016).

La “falacia de la probabilidad inversa” es la falsa creencia de que el valor de *p* indica la probabilidad de que la hipótesis nula (H_0) es cierta, dado ciertos datos ($Pr(H_0|Datos)$). Esto significa confundir la probabilidad del resultado, asumiendo que la hipótesis nula es verdadera, con la probabilidad de que la hipótesis nula sea verdadera, dados ciertos datos (Kline, 2013; Wasserstein y Lazar, 2016).

La “falacia de la replicación” vincula el valor de *p* con el grado de replicabilidad del resultado de un estudio. Supone creer erróneamente que el valor de *p* indica el grado de replicabilidad del resultado y su complemento, $1-p$, a menudo se interpreta como indicación de la probabilidad exacta de replicación (Carver, 1978; Nickerson, 2000).

La “falacia del tamaño del efecto” relaciona la significación estadística con la magnitud del efecto detectado. En concreto, supone creer erróneamente que el valor de *p* proporciona información directa sobre el tamaño del efecto (Carver, 1978). Es decir, que cuanto más pequeño es el valor de *p* más grandes son los tamaños del efecto. Sin embargo, el valor de *p* no informa sobre la magnitud de un efecto. Éste sólo puede ser determinado mediante la estimación directa de su valor con

los estadísticos apropiados y su intervalo de confianza (APA, 2010; Cumming, 2012; Cumming, Fidler, Kalinowski, y Lai, 2012; Author, 2004; Kline, 2013; Wasserstein y Lazar, 2016).

La “falacia de la significación clínica o práctica” es la falsa creencia de que el valor de p indica la importancia de los hallazgos (Nickerson, 2000; Wasserstein y Lazar, 2016). De esta manera, un efecto estadísticamente significativo es interpretado como un efecto importante. Sin embargo, un resultado estadísticamente significativo no indica que el resultado sea importante, de la misma manera que un resultado no estadísticamente significativo todavía podría ser importante.

Detrás de estas interpretaciones erróneas hay unas creencias y atribuciones sobre el significado de los resultados. Por ello, es necesario comprender el razonamiento estadístico o la forma de razonar con ideas estadísticas y dar sentido a la información estadística que realizan los psicólogos profesionales (Garfield, 2002; Leek, 2014).

OBJETIVO

El objetivo de este trabajo fue detectar los errores de razonamiento estadístico que los psicólogos profesionales españoles cometen cuando se les presentan los resultados de una prueba de inferencia estadística, dado que, su visión e interpretación de los hallazgos podría afectar a sus decisiones profesionales y poner en peligro la calidad de las intervenciones psicológicas.

Con este fin, se analizaron dos cuestiones: la primera fue la extensión de los errores más comunes de interpretación con respecto al valor de p y la segunda fue el grado en que se interpretan correctamente los valores de p .

METODO

Participantes

Se utilizó un muestreo no probabilístico. La muestra inicial estuvo formada por 113 psicólogos españoles. El 68.1% de los participantes tenía como rol principal el ejercicio profesional de la psicología, el 28.3% el rol principal de psicólogo académico, el 0.9% de investigador y el 2.7% otro rol. Puesto que el objetivo del estudio era analizar las barreras hacia una Práctica profesional basada en la Evidencia, se eliminaron los sujetos cuyo rol principal no fuese el ejercicio profesional de la Psicología.

La muestra final estuvo formada por 77 psicólogos profesionales españoles, con una edad media de 41.44 años (DT = 9.42, mínimo = 25, máximo = 64). El 31.2 fueron hombres (68.8 % mujeres). El 27.3% de los participantes fue licenciado/graduado, el 10.4% postgraduado, el 36.4% magister y el 26% doctor. La media de años como miembro del Colegio Oficial de Psicólogos de España fue de 13.73 (DT = 9.30, mínimo = 0, máximo = 33). Finalmente, el 87% de los participantes manifestaron utilizar artículos científicos en su toma de decisiones en la práctica profesional.

Finalmente, no hubo diferencias estadísticamente significativas entre hombres y mujeres en el número de años como psicólogo profesional ($F(1, 35.21) = -1.91, p = .176, d = 0.24, 95\% \text{ CI } [-0.24, 0.73]$).

Instrumento

La encuesta on-line estuvo compuesta por dos secciones. En primer lugar, la encuesta incluyó preguntas relacionadas con la información sobre: sexo, edad, nivel educativo, años de experiencia como psicólogo profesional, entorno clínico (público o privado), uso de artículos científicos para tomar decisiones en la práctica profesional (sí/no).

La segunda sección incluyó la encuesta sobre interpretaciones del valor p de Authors (2015). Este instrumento está compuesto por 10 ítems con una escala de respuesta dicotómica (verdadero

o falso) pensados para detectar interpretaciones erróneas sobre el valor p de probabilidad asociadas a las pruebas de inferencia estadística y su interpretación correcta. Las cuestiones se plantearon con el siguiente argumento:

“Supongamos que un artículo de investigación señala un valor de $p = .001$ en el apartado de resultados ($\alpha = .05$). Seleccione las afirmaciones verdaderas”.

A.-Falacia de la Probabilidad inversa:

1. Se ha probado que la hipótesis nula es verdadera.
2. Se ha probado que la hipótesis nula es falsa.
3. Se ha determinado la probabilidad de la hipótesis nula ($p = .001$).
4. Se ha deducido la probabilidad de la hipótesis experimental ($p = .001$).
5. La probabilidad de que la hipótesis nula sea verdadera, dados los datos obtenidos, es de $.001$.

B.-Falacia del tamaño del efecto:

6. El valor $p = .001$ confirma de forma directa que el tamaño del efecto ha sido grande.

C.- Falacia de la significación clínica o práctica.

7. Obtener un resultado estadísticamente significativo indica que el efecto detectado es importante.

D.- Interpretación correcta y decisión adoptada:

8. Se conoce la probabilidad del resultado de la prueba estadística, asumiendo que la hipótesis nula es cierta.

9. Dado que $p = .001$ entonces el resultado obtenido permite concluir que las diferencias no se deben al azar.

Procedimiento

Se realizó un estudio transversal mediante encuesta *on-line*. Para ello, se envió un e-mail a los Colegios Oficiales de Psicólogos invitándoles a participar en la encuesta *on-line* sobre práctica profesional en Psicología. Los potenciales participantes fueron invitados a completar una encuesta a través del uso de un sistema CAWI. Tres semanas después se envió un mensaje de seguimiento. Se controló que no hubiera duplicidades en las respuestas. Se obtuvieron 113 encuestas correctamente cumplimentadas, de las cuales 77 fueron de psicólogos cuyo rol principal era el ejercicio profesional de la psicología. La participación en el estudio fue voluntaria y completamente anónima. La recogida de datos se llevó a cabo durante los meses de mayo a septiembre de 2015.

Análisis De Datos

Se realizaron análisis de estadísticos descriptivos (frecuencias y porcentajes). Para el cálculo de los intervalos de confianza de los porcentajes se utilizaron los métodos “score” basado en los trabajos de Newcombe (2012). Estos análisis se realizaron con el programa estadístico IBM SPSS v. 20 para Windows.

RESULTADOS

La Tabla 1 presenta el porcentaje de respuestas de los participantes que seleccionan las ocho afirmaciones falsas sobre los valores de p como afirmaciones verdaderas y seleccionan la interpretación correcta del valor de p .

Tabla 1.

Porcentaje de participantes que están de acuerdo con las afirmaciones (Intervalo de Confianza 95%)

Ítem	n	%	95% IC	
			Li	Ls
Falacia de la probabilidad inversa				
1. Se ha probado que la hipótesis nula es verdadera	15	19.48	12.18	29.69
2. Se ha probado que la hipótesis nula es falsa.	15	19.48	12.18	29.69
3. Se ha determinado la probabilidad de la hipótesis nula ($p = .001$)	8	10.39	5.36	19.18
4. Se ha deducido la probabilidad de la hipótesis experimental ($p = .001$)	8	10.39	5.36	19.18
5. La probabilidad de que la hipótesis nula sea verdadera, dados los datos obtenidos, es de .001.	15	19.48	12.18	29.69
% Participantes que han valorado correctamente las cinco afirmaciones como falsas	0	0	0	4.75
Falacia de replicación				
6. Una replicación posterior tendría una probabilidad de 0.999 (1-0.001) de ser estadísticamente significativa.	3	3.90	1.33	10.84
Falacia del tamaño del efecto				
7. El valor $p = .001$ confirma de forma directa que el tamaño del efecto ha sido grande	8	10.39	5.36	19.18
Falacia de la significación clínica o práctica				
8. Obtener un resultado estadísticamente significativo indica que el efecto detectado es importante.	28	36.36	26.51	47.52
Interpretaciones correctas y decisión adoptada por el investigador				
9. Se conoce la probabilidad del resultado de la prueba estadística, asumiendo que la hipótesis nula es cierta	6	7.79	3.62	15.98
10. Dado que $p = .001$ entonces el resultado obtenido permite concluir que las diferencias no se deben al azar	36	46.75	36.03	57.78
% Participantes que han valorado correctamente las dos afirmaciones como verdaderas	1	1.30	0.02	7

Nota. IC = intervalo de Confianza; Li= límite inferior; Ls= límite superior

En general, cabe destacar que más de un tercio de todos los participantes tuvieron problemas con la falacia de la significación clínica o práctica de los resultados. Estos participantes afirmaron que el valor de p indica la importancia de los hallazgos. En consecuencia, vinculan la significación estadística de los hallazgos con su significación o relevancia clínica o práctica.

En cuanto a la “falacia de la probabilidad inversa”, la mayoría de los psicólogos percibieron que algunas de las afirmaciones falsas sobre el valor de p eran verdaderas (como el porcentaje de participantes que no endosaron las cinco afirmaciones falsas como verdaderas).

Finalmente, la mayoría de los participantes tuvieron problemas con la interpretación probabilística del valor p . La interpretación del valor de p mejoró cuando se realizó en términos de conclusión estadística, en comparación con la interpretación probabilística del valor p . Además, solo un pequeño porcentaje de participantes valoraron correctamente las dos afirmaciones como verdaderas.

DISCUSIÓN Y CONCLUSIÓN

Nuestros resultados indican que la comprensión del valor p sigue siendo problemática entre los psicólogos profesionales españoles. Interpretar un resultado estadísticamente significativo como importante o útil, confundir el nivel de significación del alfa con la probabilidad de que la hipótesis nula sea verdadera, relacionar el valor p con el tamaño del efecto y creer que la probabilidad de replicar un resultado es $1-p$ son interpretaciones erróneas que existen entre los psicólogos profesionales españoles.

La “falacia de significación clínica o práctica” fue la interpretación errónea que se observó con mayor frecuencia. Sin embargo, un resultado estadísticamente significativo no indica que el resultado sea importante, de la misma manera que un resultado no estadísticamente significativo puede ser aun importante o relevante a nivel práctico (Nickerson, 2000). La significancia clínica se refiere al valor práctico o aplicado o importancia del efecto de una intervención. Es decir, si hay una diferencia real (e.g., genuina, palpable, práctica, notable) para los clientes o para otros con quienes interactúan en la vida cotidiana (Kazdin, 1999, 2008).

La interpretación correcta del valor de p mejoró cuando se realizó en términos de conclusión estadística, en comparación con la interpretación probabilística del valor de p . Esto puede deberse al hecho de que la interpretación probabilística requiere pensar en la significación del valor de p como probabilidad condicional y una variable aleatoria, lo que significa reflexionar sobre los procesos estadísticos implicados en el comportamiento del valor de p cuando la hipótesis nula no es rechazado. Si bien la interpretación estadística sólo se basa en la valoración del valor p , proporcionada por los resultados de los programas estadísticos, en comparación con el valor alfa α , por tanto, sólo requiere que el investigador aplique rutinariamente la regla $p < \alpha$ (Pfannkuch & Wild, 2004).

Los problemas en la comprensión del valor p pueden afectar las conclusiones que los profesionales extraen de las investigaciones psicológicas (Hoekstra, Morey, Rouder, y Wagenmakers, 2014), poniendo en peligro la calidad de los resultados de la investigación psicológica (Autor, 2011). El valor de la evidencia científica depende de la calidad de los análisis estadísticos y de su interpretación (Faulkner, Fidler, y Cumming, 2008). Estos conceptos erróneos son problemas de interpretación y no son un problema del procedimiento de la NHST en sí mismo (Leek, 2014). Detrás de estas interpretaciones erróneas existen algunas creencias y atribuciones acerca de la significación estadística de los resultados. Por lo tanto, es necesario mejorar la enseñanza de la estadística, la formación de los psicólogos y el contenido de los manuales de estadística con el fin de garantizar una formación de alta calidad a los futuros profesionales (Babione, 2010; Cumming, 2012; Kline, 2013; Haller y Krauss, 2002).

Las pruebas de significación estadística tienen un propósito y responden a algunos problemas y no a otros (Perezgonzalez, 2015). Una prueba de significancia estadística no habla de la impor-

tancia del resultado, la replicabilidad de los resultados (Carver, 1978; Verdam, Oort, y Sprangers, 2014). El valor de p nos informa de si existe un efecto o no, pero no nos indica el tamaño de dicho efecto, ni el significado clínico/práctico del mismo (Sullivan y Feinn, 2012). El tamaño del efecto sólo puede determinarse estimando directamente su valor con los estadísticos apropiados y su intervalo de confianza (Cohen, 1994; Cumming, 2012; Kline, 2013; Téllez, García, y Corral-Verdugo, 2015).

Finalmente, deben mencionarse algunas limitaciones de este estudio. La baja tasa de respuesta podría afectar la representatividad de la muestra y, por lo tanto, la generalización de los hallazgos entre los psicólogos profesionales españoles. Sin embargo, es posible que los participantes que respondieron a la encuesta se sintieran más seguros sobre sus conocimientos estadísticos que los que no respondieron. Si esto fuera así, los resultados podrían infraestimar la extensión de las falacias sobre el valor p entre los psicólogos profesionales españoles.

Además, los resultados del presente estudio coinciden con los hallazgos de estudios previos sobre conceptos erróneos del valor p en muestras de psicólogos académicos y estudiantes de grado de Psicología (Authors 2015, 2016, 2017; Falk y Greenbaum, 1995, Haller y Krauss, 2002, Kühberger y cols., 2015, Monterde-i-Bort y cols., 2010; Oakes, 1986).

Todo esto lleva a señalar la necesidad de dotar de una formación metodológica adecuada a los profesionales de la Psicología, la cual podría mejorar su práctica profesional. La PBE requiere que los profesionales evalúen críticamente los resultados de las investigaciones psicológicas (Daset y Cracco, 2013). Para poder hacerlo, se necesita formación en conceptos estadísticos, metodología de diseño de investigación y resultados de pruebas de inferencia estadística.

AGRADECIMIENTOS

Este artículo se basa en la Tesis Doctoral con Mención Internacional defendida por Author (2016)....

REFERENCIAS

- Autor 2003
 Autor 2004
 Autor, 2010
 Autor 2014
 Autor 2015
 Autor 2016
 Autor 2017
 American Psychological Association (2005). *Policy Statement on Evidence-Based Practice in Psychology*. Washington, DC: Author.
 American Psychological Association (2006). Evidence-based practice in psychology: APA Presidential Task Force on evidence-based practice. *American Psychologist*, *61*, 271–285. doi:10.1037/0003-066X.61.4.271
 American Psychological Association (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
 Babione, J. M. (2010). Evidence-Based Practice in Psychology: An ethical framework for graduate education, clinical training, and maintaining professional competence. *Ethics & Behavior*, *20*, 443-453. doi: 10.1080/10508422.2010.521446
 Beyth-Maron, R., Fidler, F., y Cumming, G. (2008). Statistical cognition: Towards evidence-based practice in statistics and statistics education. *Statistics Education Research Journal*, *7*, 20-39.
 Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*, 378-399.

- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003. doi: 10.1037/0003-066X.49.12.997
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.
- Cumming, G., Fidler, F., Kalinowski, P. y Lai, J. (2012). The statistical recommendations of the American Psychological Association publication manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, 64, 138-146. doi: 10.1111/j.1742-9536.2011.00037.x
- Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., ..., y Wilson, S. (2007). Statistical reform in Psychology: Is anything changing? *Psychological Science*, 18, 230-232. doi: 10.1111/j.1467-9280.2007.01881.x
- Daset, L. R., y Cracco, C. (2013). Psicología Basada en la Evidencia: algunas cuestiones básicas y una aproximación a través de una revisión bibliográfica. *Ciencias Psicológicas*; 7, 209 – 220.
- Falk, R., y Greenbaum, C. W. (1995). Significance tests die hard: the amazing persistence of a probabilistic misconception. *Theory y Psychology*, 5, 75-98. doi: 10.1177/0959354395051004
- Faulkner, C., Fidler, F., y Cumming, G. (2008). The value of RCT evidence depends on the quality of statistical analysis. *Behavior Research and Therapy*, 46, 270-281. doi: 10.1016/j.brat.2007.12.001
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of statistic education*, 10. Disponible en www.amstat.org/publications/jse/v10n3/garfield.html
- Haller, H., y Krauss, S. (2002). Misinterpretations of significance: a problem students share with their teachers? *Methods of Psychological Research Online [On-line serial]*, 7, 120. Recuperado de: <http://www.metheval.uni-jena.de/lehre/0405-ws/evaluationuebung/haller.pdf>
- Harrison, J., Thompson, B., y Vannest, K. J. (2009). Interpreting the evidence for effective interventions to increase the academic performance of students with ADHD: Relevance of the statistical significance controversy. *Review of Educational Research*, 79, 740-775. doi: 10.3102/0034654309331516
- Hoekstra, R., Morey, R.D., Rouder, J.N., y Wagenmakers, E. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21, 1157-1164. doi: 10.3758/s13423-013-0572-3
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 332-339. doi: 10.1037/0022-006X.67.3.332
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63, 146-159. doi: 10.1037/0003-066X.63.3.146
- Kline, R. B. (2013). *Beyond significance testing: Statistic reform in the behavioral sciences*. Washington, DC: American Psychological Association
- Kühberger, A., Fritz, A., Lermer, E., y Scherndl, T. (2015). The significance fallacy in inferential statistics. *BMC Research Notes*, 17, 8, 84. doi: 10.1186/s13104-015-1020-4.
- Leek, J. (2014). On the scalability of statistical procedures: Why the p-value bashers just don't get it. Simply Statistics Blog, Disponible en: <http://simplystatistics.org/2014/02/14/on-the-scalability-of-statistical-procedures-why-the-p-value-bashers-just-dont-get-it/>
- Monterde-i-Bort, H., Frias-Navarro, D., & Pascual-Llobell, J. (2010). Uses and abuses of statistical significance tests and other statistical resources: A comparative study. *European Journal of Psychology of Education*, 25, 429-447. doi: 10.1007/s10212-010-0021-x
- Newcombe, R.G. (2012). *Confidence intervals for proportions and related measures of effect size*. Boca Raton, FL: CRC Press

- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301. doi: 10.1037/1082-989X.S.2.241
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. Chichester: John Wiley & Sons.
- Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology*, 6, 223. doi: 10.3389/fpsyg.2015.000223
- Pfannkuch, M., y Wild, C. (2004). Towards an understanding of statistical thinking. En D. Ben-Zvi, y J. Garfield, (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-45). Dordrecht: Kluwer Academic Publishers.
- Sackett, D.L., Straus, S. L., Richardson, W.S., Rosenberg, W.M.C. y Haynes, R. B. (2000). *Evidence Based Medicine. How to practice y teach EBM (3rd Edition)*. Edinburgh, England & New York (NY): Churchill Livingstone.
- Sánchez-Meca, J., Boruch, R.F., Petrosino, A., y Rosa-Alcázar, A.I. (2002). La Colaboración Campbell y la Práctica basada en la Evidencia. *Papeles del Psicólogo*, 83, 44-48. Disponible en: <http://www.papelesdelpsicologo.es/resumen?pii=896>
- Sullivan, G. M., y Feinn, R. (2012). Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, 4, 279–282. doi: 10.4300/JGME-D-12-00156.1
- Téllez, A., García, C. H., y Corral-Verdugo, V. (2015). Effect size, confidence intervals and statistical power in psychological research. *Psychology in Russia: State of the Art*, 8, 27-46. doi: 10.11621/pir.2015.0303
- Vázquez, C., y Nieto, M. (2003). Psicología (clínica) basada en la evidencia (PBE): una revisión conceptual y metodológica. En J. L. Romero (Ed), *Psicópolis: Paradigmas actuales y alternativos en la psicología contemporánea*. Barcelona. Paidós.
- Verdam, M. G. E., Oort, F. J., & Sprangers, M. A. G. (2014). Significance, truth and proof of p values: reminders about common misconceptions regarding null hypothesis significance testing. *Quality of Life Research*, 23, 5–7. doi: 10.1007/s11136-013-0437-2
- Wasserstein, R. L., y Lazar, N. A. (2016). The ASA's Statement on p-values: Context, process, and purpose. *The American Statistician*, 70, 129-133. doi: 10.1080/00031305.2016.1154108

