


Editorial

# El Valor $p$ No es Suficiente: Importancia de Reportar los Intervalos de Confianza y el Tamaño del Efecto en la Investigación Psicológica

The  $p$ -Value Is Not Enough: Importance of Reporting Confidence Interval and Effect Size in Psychological Research

Juan Aníbal González Rivera <sup>1</sup> & Adam Rosario-Rodríguez <sup>2</sup>

1 Ponce Health Sciences University, San Juan, Puerto Rico.  <https://orcid.org/0000-0003-0622-8308>

2 Universidad Albizu, Recinto de San Juan, Puerto Rico.  <https://orcid.org/0000-0003-1603-5040>

\* Correspondencia: [jagonzalez@psm.edu](mailto:jagonzalez@psm.edu)

[WWW.REVISTACARIBENADEPSICOLOGIA.COM](http://WWW.REVISTACARIBENADEPSICOLOGIA.COM)

## Citar como:

González-Rivera, J. A., & Rosario-Rodríguez, A. (2024). El valor  $p$  no es suficiente: Importancia de reportar los intervalos de confianza y el tamaño del efecto en la investigación psicológica. *Revista Caribeña de Psicología*, 8, e8787. <https://doi.org/10.37226/rcp.v8i1.8787>

## Editorial Volumen 8, 2024

### Introducción

En el ámbito de la investigación psicológica, el valor  $p$  ha sido durante mucho tiempo el estándar de oro para determinar la significancia estadística de los resultados (Funder & Ozer, 2019). Este enfoque ha dominado el campo, dictando en gran medida la interpretación y la aceptación de los hallazgos en la comunidad científica. Sin embargo, dos de los estadísticos investigadores más influyente de las últimas décadas declararon lo siguiente:

*“La significación estadística es lo menos interesante de los resultados. Hay que describir los resultados en términos de medidas de magnitud: no sólo si un tratamiento afecta a las personas, sino cuánto las afecta”.*

Gene V. Glass (citado en Kline, 2004)

*“El producto principal de una investigación es una o más medidas del tamaño del efecto, no los valores  $p$ ”.*

Jacob Cohen (Cohen, 1990)

Recientemente, una creciente corriente de críticas y cuestionamientos ha emergido en torno a la

dependencia exclusiva del valor  $p$ . Argumentos convincentes sugieren que este enfoque, por sí solo, es insuficiente para capturar la complejidad y la relevancia práctica de los resultados de la investigación. Por ello, en la actualidad, se le está prestando mayor atención e importancia a los tamaños del efecto (e.g., Andradre, 2019; Cumming, 2012) y a los intervalos de confianza (Das, 2022; Samargandi et al., 2018). Por esta razón, con este editorial pretendemos abogar por la inclusión y el énfasis en el intervalo de confianza junto con el tamaño del efecto en los manuscritos sometidos a nuestra revista.

### Valor $p$

El valor  $p$  se define como la probabilidad de obtener un resultado al menos tan extremo como el observado, asumiendo que la hipótesis nula es verdadera. En términos más sencillos, es una medida que indica la fuerza de la evidencia contra la hipótesis nula. Tradicionalmente, un valor  $p$  menor a 0.05 se ha considerado como un umbral para declarar la significancia estadística (Dahiru, 2008). Sin embargo, este umbral

no es una frontera mágica entre lo significativo y lo no significativo, sino más bien una convención que ha prevalecido en el tiempo.

Una de las principales limitaciones del valor  $p$  es su frecuente malinterpretación. Un valor  $p$  bajo no necesariamente indica un efecto grande o importante, sino simplemente que la evidencia contra la hipótesis nula es fuerte bajo los supuestos del modelo estadístico utilizado. Además, el valor  $p$  puede ser influenciado por el tamaño de la muestra: en muestras grandes, incluso efectos trivialmente pequeños pueden resultar en valores  $p$  bajos (Andradre, 2019). Esta situación plantea el riesgo de sobrevalorar los resultados. Además, el valor  $p$  no ofrece información sobre la magnitud del efecto observado, lo que es crucial para entender la relevancia práctica de los hallazgos.

Dahiru (2008) plantea que existen algunas falacias relacionadas a la comprobación de hipótesis y el uso del valor  $p$ :

1. El uso de un alfa de 0.05 es una norma con una base objetiva. Incorrecto; la distinción entre resultados “significativos” y “no significativos” no es tajante. Lo que existe es una acumulación progresiva de evidencia en contra de la hipótesis nula a medida que el valor  $p$  disminuye. Por ejemplo, un valor  $p$  de 0.02 proporciona evidencia más sólida que un valor  $p$  de 0.04.
2. Un valor  $p$  pequeño indica efectos grandes. Falso; el valor  $p$  no dice nada sobre el tamaño de un efecto.
3. La significancia estadística implica importancia clínica o práctica. Falso; la significancia estadística no necesariamente refleja la relevancia clínica de una relación. Existe una marcada distinción entre la significancia estadística y la importancia clínica. Bajo el criterio estadístico estándar de  $\alpha = 0.05$ , hay una expectativa de que, en 1 de cada 20 comparaciones donde la hipótesis nula es verdadera, obtendremos un resultado con  $p < 0.05$ . Esta situación ilustra una de las muchas falacias asociadas con las pruebas de hipótesis. Resulta preocupante observar en publicaciones científicas cómo las pruebas de significancia a menudo se malinterpretan o se utilizan inadecuadamente,

reduciendo su valor a meras pruebas de “insignificancia”.

Ilustraremos la cuestión de la relevancia práctica o clínica a través de un ejemplo extraído de una investigación propia. En 2019, junto con colegas, publicamos un estudio titulado “¿Realmente son los creyentes más felices que los ateos? Medidas de bienestar en una muestra de ateos y creyentes en Puerto Rico” (González-Rivera et al., 2019). El propósito de este estudio fue examinar las posibles diferencias en los niveles de satisfacción vital y florecimiento psicológico entre individuos que se identifican como creyentes en Dios y aquellos que se autoidentifican como ateos.

En nuestra investigación, la media de satisfacción vital en el grupo de creyentes en Dios fue de 28.13 ( $M_{\text{rango}} = 429.87$ ), en contraste con 27.06 ( $M_{\text{rango}} = 391.71$ ) en el grupo ateo, indicando una diferencia marginal de 1.07 puntos en la media y de 38.16 en la media de los rangos. Utilizando la prueba no paramétrica U de Mann-Whitney, se identificaron diferencias estadísticamente significativas entre los grupos,  $U = 76.412.50$ ,  $Z = -2.311$ ,  $p = 0.02$ ,  $r = 0.08$ . Estos hallazgos apuntan a una ligera superioridad en términos de satisfacción vital entre los creyentes comparados con los ateos. No obstante, es crucial destacar que el tamaño del efecto encontrado fue pequeño.

Para una interpretación rigurosa de estos resultados, se deben considerar varias preguntas clave: (1) ¿Representa una diferencia de un punto, en una escala de 5 a 35, un cambio significativo en el bienestar real de las personas? (2) Si la satisfacción vital media de ambos grupos se categoriza como alta, ¿existen diferencias prácticas o clínicas de relevancia entre creyentes y ateos? (3) ¿Implica esta mínima diferencia que los creyentes experimentan una vida psicológicamente más próspera o con mayor sentido que los ateos? La respuesta a estas interrogantes es definitivamente que no. Por consiguiente, una interpretación adecuada y reflexiva de estos datos sería: “A pesar de la existencia de una diferencia estadísticamente significativa de un punto en la media de satisfacción vital entre los grupos, esta variación no parece poseer una relevancia práctica considerable. Esto se debe a que ambos grupos alcanzan niveles altos de satisfacción y el tamaño del efecto es pequeño”.

En síntesis, aunque el valor  $p$  sigue siendo una herramienta útil en el análisis estadístico, es crucial

reconocer sus limitaciones y complementarlo con otras medidas, como el intervalo de confianza y el tamaño del efecto, para obtener una comprensión más rica y precisa de los resultados de la investigación psicológica. El uso exclusivo del valor  $p$  puede llevar a conclusiones erróneas o a una interpretación inadecuada de los datos.

Hay que tener en cuenta que el valor  $p$  añade poco o nada al intervalo de confianza del 95%. En este sentido, se promueve el uso de estimaciones del tamaño del efecto junto con los intervalos de confianza, y se podría complementar con valores  $p$  si fuera necesario (Tijseen, 2021).

### Intervalos de Confianza

Los intervalos de confianza presentan la variabilidad observada entre la medida obtenida en el estudio y la medida real de la población (Candia & Caiozzi, 2005). Estos se presentan en un rango que va desde el límite inferior (2.5%) hasta el límite superior (97.5%). Este rango de valores sigue una distribución normal donde se encuentra el valor real del valor estimado con una alta probabilidad. El consenso establecido en la comunidad científica para esa “alta probabilidad” es de un 95%. Así que, un intervalo de confianza del 95% lo que nos indica es que, dentro del rango de valores, tenemos un 95% de confianza de que se encuentre el valor real poblacional (Candia & Caiozzi, 2005).

Desde un punto de vista práctico e informativo, los intervalos de confianza dan más información de lo que podría dar el valor  $p$ , y no se ve afectado por el tamaño de la muestra. Asimismo, una de las ventajas de los intervalos de confianza es que se pueden calcular para cualquier análisis inferencial ya sean de comparaciones de grupos, de relaciones entre variables, modelos predictivos, y hasta para las medidas de consistencia interna y los tamaños del efecto (Das, 2019).

Los intervalos de confianza son relativamente simples de interpretar. Si ambos límites (inferior y superior) son positivos o negativos es indicativo de que tenemos significancia. Para pruebas de diferencias de grupos o relaciones entre variables, debemos observar si los límites no incluyen el 0 entre sus posibilidades para lograr significancia. Una manera de tener clara la presencia del 0 es cuando el límite inferior es negativo y el superior es positivo, en ese escenario no

hay evidencia de significancia. Por otro lado, para estudios en donde se utilizan el *Odds Ratio* o análisis de riesgo se debe observar si el rango entre los límites inferior y superior no incluye el 1, porque en caso de que lo incluya, es evidencia de que no hay significancia (Adedokun, 2008).

Los intervalos de confianza son una forma de estimar la incertidumbre en las estimaciones estadísticas. Como científicos debemos aprender a aceptar, abrazar y reconocer la incertidumbre. El reportar los intervalos de confianza junto con el tamaño del efecto brindará más información de nuestros hallazgos de lo que nos daría solamente el valor  $p$ . Además, debemos tener presente el contexto de nuestros datos para mejorar las interpretaciones de nuestros hallazgos. Asimismo, no podemos olvidar que el contexto es tan importante como la propia calidad de los datos para lograr estimaciones e inferencias con mayor estimación.

### Tamaño de Efecto

El tamaño del efecto es una medida cuantitativa de la magnitud de un fenómeno (Ferguson, 2009). A diferencia del valor  $p$ , que solo indica si un resultado es estadísticamente significativo, el tamaño del efecto informa sobre la magnitud de la diferencia o la relación observada. Hay diversas formas de calcular el tamaño del efecto, como la  $d$  de Cohen, que mide la diferencia entre dos medias en términos de desviaciones estándar, o la  $r$  de Pearson, que evalúa la fuerza de una correlación (Funder & Ozer, 2019).

Reportar el tamaño del efecto tiene varias ventajas. Primero, proporciona una comprensión más clara de la relevancia práctica de los resultados (Ferguson, 2009). Como ya hemos mencionado, un estudio podría tener un valor  $p$  significativo, pero si el tamaño del efecto es pequeño, el resultado puede no ser relevante en un contexto práctico. Segundo, el tamaño del efecto permite comparar los resultados a través de estudios, incluso cuando estos utilizan diferentes escalas o medidas. Esto es particularmente útil en metaanálisis y revisiones sistemáticas. Tercero, ayuda a evitar malinterpretaciones comunes del valor  $p$ , como la falsa creencia de que un valor  $p$  bajo implica automáticamente un efecto grande o importante (Dahiru, 2008).

Numerosos estudios han demostrado cómo el tamaño del efecto puede revelar información que el

valor  $p$  solo no puede. Por ejemplo, en investigaciones donde se comparan diferentes tratamientos psicológicos, el valor  $p$  puede indicar que hay una diferencia significativa entre los tratamientos, pero el tamaño del efecto revelará cuán grande y clínicamente relevante es esa diferencia. Por esta razón, muchos investigadores llegan a conclusiones distintas de los efectos de un mismo tratamiento (Ferguson, 2009). Esto es crucial para los profesionales que deben tomar decisiones basadas en la efectividad relativa de diferentes intervenciones.

En cuanto al uso del tamaño del efecto en la investigación psicológica, debemos mencionar cinco puntos importantes:

1. Los investigadores que publican en español no suelen reportar el tamaño del efecto. Esta es la razón principal por la que hemos decidido escribir esta editorial.
2. Los tamaños de los efectos suelen ser muy pequeños (Rosnow & Rosenthal, 2003). Esto es un problema a la hora de interpretar resultados.
3. No existe un consenso claro respecto a la magnitud del efecto requerido para determinar una significación práctica.
4. A los tamaños del efecto se le puede calcular el intervalo de confianza. Esto da información útil del propio tamaño del efecto y el rango de posibilidades en donde se mueve el efecto.
5. Las interpretaciones clásicas del tamaño del efecto, como las que sugiere Cohen (1992), son un mínimo sugerido y no una garantía de que los tamaños del efecto obtenidos sean significativos en la práctica. La Tabla 1 presenta las sugerencias de Ferguson (2009) para la interpretación del tamaño del efecto. Sus guías son muy interesantes y útiles para nuestro campo ya que recomienda un tamaño mínimo del efecto que represente un efecto “prácticamente” significativo para los datos de las ciencias sociales.

La literatura sugiere que los tamaños del efecto se dividen en cuatro categorías principales (Ferguson, 2009; Kline, 2004; Vacha-Haase & Thompson, 2004):

1. **Índices de Diferencia entre Grupos:** Estos índices se enfocan en cuantificar la magnitud de las diferencias observadas entre dos o más grupos. Un ejemplo prominente de este tipo es la  $d$  de Cohen, que mide la diferencia entre medias en términos de desviaciones estándar. También tenemos la alternativa delta ( $\Delta$ ), que supone un grupo control; y la  $g$  de Hedges (1981), que es más estandarizada que las anteriores y representa la mejor opción para los diseños del ANOVA (Ferguson, 2009).
2. **Índices de Fuerza de Asociación:** Estos indicadores evalúan la magnitud de la relación, es decir, cuánta varianza es compartida entre dos o más variables. Un ejemplo clásico es la correlación  $r$  de Pearson, que mide el grado y la dirección de la asociación entre variables. También contamos con las alternativas  $R$ , Spearman rho ( $\rho$ ),  $r$  parcial,  $V$  de Cramér ( $\varphi$ ), coeficiente de regresión  $\beta$ , y  $r_h$ .
3. **Estimaciones Corregidas:** Estas medidas ajustan los resultados para compensar posibles errores de muestreo, especialmente relevantes en muestras de menor tamaño. Un ejemplo es la  $R^2$  ajustada, que proporciona una estimación más precisa de la varianza explicada en modelos de regresión. También tenemos las alternativas  $r^2$ , Eta-cuadrada ( $\eta^2$ ), Omega cuadrada ( $\omega^2$ ) y Epsilon cuadrada ( $\epsilon^2$ ).
4. **Estimaciones de Riesgo:** Estas medidas son cruciales en la investigación en salud y se utilizan para comparar el riesgo relativo de un evento específico entre dos o más grupos. Incluyen estadísticas como el *Riesgo Relativo* (RR) y el *Odds Ratio* (OR), ambas fundamentales para evaluar y comparar riesgos en diferentes condiciones o tratamientos.

En conclusión, los intervalos de confianza y el tamaño del efecto son unas herramientas estadísticas esenciales en la investigación psicológica. Proporciona una visión más completa y matizada de los resultados de investigación y deben ser reportados y considerados junto con el valor  $p$  para una interpretación adecuada y completa de los hallazgos. Al hacerlo, los investigadores pueden ofrecer una perspectiva más rica y práctica sobre sus estudios, mejorando

la calidad y la aplicabilidad de la investigación en psicología.

**Tabla 1**

*Sugerencias de interpretación del tamaño del efecto para investigaciones en psicología.*

Índices	TEMR	Efecto Moderado	Efecto Fuerte
$d, \Delta, g$	.41	1.15	2.70
$r, R, \rho, r$ parcial, $\varphi, \beta, r_i$	.2	.5	.8
$r^2, R^2, R^2$ ajustada, $\eta^2, \omega^2, \epsilon^2$	.04	.25	.64
RR, OR	2.0	3.0	4.0

*Nota.* Basado en las recomendaciones de Ferguson (2009). TEMR = tamaño del efecto mínimo recomendado en psicología.

### Recomendaciones Finales

A medida que la comunidad de investigación en psicología se esfuerza por alcanzar niveles más altos de precisión y profundidad en sus estudios, se hace cada vez más evidente la necesidad de adoptar metodologías que reflejen estos avances. Para ello, ofreceremos algunas recomendaciones finales, especialmente en lo que respecta al reporte de los intervalos de confianza y el tamaño del efecto, y su integración con el valor  $p$ , con el objetivo de mejorar sustancialmente la calidad de la investigación en psicología.

Una recomendación fundamental es la incorporación de los intervalos de confianza junto con el tamaño del efecto en los reportes de investigación. Para lograr esto, los investigadores deben calcular el intervalo de confianza para sus estimaciones y seleccionar la medida de tamaño del efecto más adecuada en función del tipo de análisis estadístico empleado. Por ejemplo, la  $d$  de Cohen es apropiada para comparar medias, mientras que la  $r$  de Pearson es ideal para evaluar correlaciones. Es crucial que los resultados del tamaño del efecto se presenten de forma clara y comprensible, evitando una jerga excesivamente técnica y explicando su significado en el contexto específico del estudio. Además, siempre que sea posible, los investigadores deben contextualizar sus hallazgos comparándolos con estudios previos o puntos de referencia establecidos, proporcionando así una referencia para interpretar la relevancia del efecto observado.

Otra consideración importante es la combinación del valor  $p$  con el intervalo de confianza y el tamaño del efecto para una comprensión más integral de los resultados. Mientras que el valor  $p$  indica si un resultado es estadísticamente significativo, el intervalo de confianza da una idea amplia del rango de posibilidades donde se sitúa el estimado en relación con el valor real de la población, y el tamaño del efecto aporta información sobre la magnitud práctica de este efecto. Esta interpretación triangular permite a los investigadores proporcionar una visión más completa de sus hallazgos. Además, ayuda a evitar la sobreinterpretación de resultados que son estadísticamente significativos, pero tienen un tamaño del efecto pequeño, lo que es crucial para una comprensión matizada y realista de los datos. Al enfocarse tanto en el valor  $p$ , como en los intervalos de confianza y en el tamaño del efecto, los investigadores pueden evaluar de manera más efectiva la importancia práctica y teórica de sus hallazgos, algo esencial en las investigaciones psicológicas y de salud en general.

El compromiso con estas prácticas de reporte no solo enriquecerá la calidad de la investigación psicológica, sino que también fomentará una mayor transparencia y relevancia en el campo. Se alienta a los investigadores a adoptar un enfoque holístico en sus análisis estadísticos, considerando tanto la significancia estadística como la importancia práctica de sus hallazgos, para impulsar el conocimiento y la aplicación efectiva de la psicología.

### Novedades de la Revista

#### *Reportar los Intervalos de Confianza y el Tamaño del Efecto*

En la Revista Caribeña de Psicología deseamos comunicar un cambio significativo en nuestras políticas editoriales que entrarán en vigor a partir de enero de 2024. En un esfuerzo por mejorar la rigurosidad y transparencia en la investigación psicológica, hemos decidido hacer obligatorio el reporte de los intervalos de confianza y los tamaños del efecto en todas las investigaciones científicas que se publiquen en nuestra revista. Este cambio refleja nuestro compromiso continuo con la excelencia y la integridad en la investigación científica.

La decisión de exigir el reporte de los intervalos de confianza y los tamaños del efecto surge del

reconocimiento de que estas medidas proporcionan una información crucial que va más allá de la mera significación estadística. El tamaño del efecto ofrece una comprensión más profunda de la magnitud de los hallazgos y su relevancia práctica, facilitando así una evaluación más completa y matizada de los resultados de investigación. Esta política no solo alinearán nuestra revista con las mejores prácticas internacionales en la investigación psicológica, sino que también ayudará a los investigadores y lectores a interpretar los resultados de manera más efectiva.

Instamos a todos los autores que deseen contribuir a nuestra revista a prepararse para esta nueva exigencia. En los próximos meses, proporcionaremos directrices detalladas y recursos educativos para asistir en la implementación de esta política. Confiamos en que esta iniciativa enriquecerá la calidad de las investigaciones publicadas y reforzará la utilidad práctica de los estudios para la comunidad psicológica y académica.

### Sección de Debates

La Revista Caribeña de Psicología y su editor en jefe, reconociendo la importancia crítica del debate en el avance de la psicología, han introducido una sección dedicada exclusivamente al debate titulada "Debates, Reacciones y Controversias Teóricas". Esta iniciativa busca abordar los desafíos actuales en la disciplina y fomentar un entorno académico más dinámico y colaborativo. El principal objetivo de esta sección es proporcionar una plataforma para el intercambio abierto y riguroso de ideas, fomentando la discusión crítica y el cuestionamiento constructivo.

La sección "Debates, Reacciones y Controversias Teóricas" invita a presentaciones que aborden desde críticas constructivas de investigaciones recientes hasta discusiones sobre temas emergentes y controvertidos en la psicología. La estructura será flexible, permitiendo diversos formatos como artículos de opinión, respuestas a publicaciones previas y paneles de discusión. Los temas serán variados, cubriendo áreas como la psicología clínica, social, educativa y organizacional, con un énfasis particular en la inclusión de perspectivas multidisciplinares y culturales. La revisión por pares en esta sección será conducida bajo principios de rigor académico, imparcialidad y apertura. Se promoverá un ambiente de respeto y profesionalismo.

### Convocatoria Abierta

La Revista Caribeña de Psicología es la revista oficial y científica de la Fundación para el Desarrollo de la Psicología en el Caribe (FUNDEPSIC), entidad sin fines de lucro, que tiene como objetivo fundamental promover el avance y desarrollo de la Psicología en el Caribe y América Latina. La revista se publica de forma continua y los manuscritos sometidos serán publicados inmediatamente sean aceptados. Por tal razón, se mantendrá una convocatoria permanente para la recepción de artículos.

Los artículos aceptados serán publicados inmediatamente en el volumen correspondiente. El tiempo promedio de revisión de un manuscrito es entre 120 y 180 días desde que es recibido. Recordamos que los artículos deben someterse en formato Word y en estilo APA 7ma edición. Todo artículo debe ser sometido a través de nuestra página web y siguiendo al pie de la letra las directrices para los autores. Les invitamos a seguir patrocinando la revista, tanto a los lectores asiduos como a los autores.

### REFERENCIAS

- Adedokun, B. O. (2008). P-value and confidence interval – Facts and farces. *Annals of Ibadan Postgraduate Medicine*, 6(1), 33-34. <https://doi.org/10.4314%2Faipm.v6i1.64041>
- Andrade, C. (2019). The  $P$  Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian journal of psychological medicine*, 41(3), 210–215. [https://doi.org/10.4103/IJPSYM.IJPSYM\\_193\\_19](https://doi.org/10.4103/IJPSYM.IJPSYM_193_19)
- Candia, R. B., & Caiozzi, G. A. (2005). Intervalos de confianza. *Revista Médica de Chile*, 133(9), 1111-1115. <https://dx.doi.org/10.4067/S0034-98872005000900017>
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304–1312. <https://doi.org/10.1037/0003-066X.45.12.1304>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge/Taylor & Francis Group.
- Dahiru T. (2008). P - value, a true test of statistical significance? A cautionary note. *Annals of Ibadan postgraduate medicine*, 6(1), 21–26. <https://doi.org/10.4314/aipm.v6i1.64038>
- Das, S. K. (2019). Confidence interval is more informative than p-value in research. *International Journal of Engineering Applied Sciences and Technology*, 4(6), 278-282. <https://www.ijeast.com/papers/278-282,Tesma406,IJEAST.pdf>

- Das, V. (2022). Why you should prefer confidence interval over p-value: Communicating results of your statistical analysis. *Medium*. <https://towardsdatascience.com/why-you-should-prefer-confidence-interval-over-p-value-e32293bd174c>
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538. <https://doi.org/10.1037/a0015808>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- González-Rivera, J. A., Rosario-Rodríguez, A., Rodríguez-Ramos, E., Hernández-Gato, I., & Torres-Báez, L. M. (2019). Are believers happier than atheists? Well-being measures in a sample of atheists and believers in Puerto Rico. *Interacciones*, 5(1), 51-59. <http://dx.doi.org/10.24016/2019.v5n1.160>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.2307/1164588>
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. American Psychological Association. <https://doi.org/10.1037/10693-000>
- Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57(3), 221–237. <https://doi.org/10.1037/h0087427>
- Samargandi, O. A., Al-Taha, M., Moran, K., Al Youha, S., & Bezuhly, M. (2018). Why the p value alone is not enough: The need for confidence intervals in plastic surgery research. *Plastic and Reconstructive Surgery*, 141(1), 152e–162e. <https://doi.org/10.1097/PRS.00000000000003960>
- Tijssen, J. G. P. (2021). More confidence interval and fewer p values: A positive trend? *Journal of the American College of Cardiology*, 77(22), 1562-1563. <https://www.jacc.org/doi/10.1016/j.jacc.2021.02.004>
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473–481. <https://doi.org/10.1037/0022-0167.51.4.473>



Obra bajo licencia de Creative Commons Atribución 4.0 Internacional (CC BY 4.0).  
© 2024 Autores.