



Espiral. Cuadernos del Profesorado
Vol. 18, Núm. 38
Septiembre 2025
ISSN 1988-7701

Cómo referenciar este artículo / How to reference this article:

Rodríguez-Sabiote, C., Vázquez, L.M., López-López, J.A., & Sánchez-Martín, M. (2025). Effect size as an alternative to statistical significance testing. A practical example with JASP. *Espiral. Cuadernos del Profesorado*, 18(38), 129-141. <https://doi.org/10.25115/ecp.v18i38.10234>

Publicado on-line: 25 de septiembre de 2025

El tamaño del efecto como alternativa a las pruebas de significación estadística. Un ejemplo práctico con JASP¹

Effect size as an alternative to statistical significance testing. A practical example with JASP

Clemente Rodríguez Sabiote ¹, Lindsay Michelle Vázquez ¹, José Antonio López-López ²
y Micaela Sánchez-Martín ²

¹ Universidad de Granada, Granada, España; ² Universidad de Murcia, Murcia, España

“La significación estadística no es lo mismo que la importancia práctica” (David A. Friedman)

Resumen

El texto aborda la crítica a la dependencia exclusiva de la significación estadística en la investigación científica y propone el uso del tamaño del efecto (TE) como una alternativa más robusta y explicativa. Se destacan dos principales debilidades de la significación estadística: su dependencia del tamaño muestral y la arbitrariedad del umbral del valor p (generalmente 0.05). Estas limitaciones pueden llevar a interpretaciones erróneas y prácticas cuestionables como el "p-hacking" o dragado de datos. El tamaño del efecto se presenta como una medida cuantitativa de la magnitud de un fenómeno, independiente del tamaño de la muestra, que facilita la comparación entre estudios y contextos diversos. Se clasifica en varias tipologías: diferencias de medias (d de Cohen, g de Hedges, Δ de Glass), correlaciones (r de Pearson, r^2), análisis de varianza (η^2 , ω^2 , f de Cohen) y razones de probabilidades o cuotas (Odds Ratio, Risk Ratio). La interpretación del TE varía según el contexto, pero guías generales sugieren que valores como 0.2, 0.5 y 0.8 en d de Cohen representan efectos pequeños, medianos y grandes, respectivamente. Finalmente, se incluye un ejemplo práctico para ilustrar la aplicación de estas medidas y como la significación estadística y los tamaños del efecto pueden llegar a conclusiones contrarias.

Palabras clave: Tamaños del efecto; significación estadística; JASP.

Abstract

The work addresses the criticism of the exclusive reliance on statistical significance (SS) in scientific research and proposes the use of effect size (ES) as a more robust and explanatory alternative. Two main weaknesses of SS are highlighted: its dependence on sample size and the arbitrariness of the p-value threshold (generally 0.05). These limitations can lead to misinterpretations and questionable practices such as "p-hacking" or data dredging. Effect size is presented as a quantitative measure of the magnitude of a phenomenon, independent of sample size, facilitating comparison between studies and diverse contexts. It is classified into various typologies: mean differences (Cohen's d, Hedges' g, Glass' Δ), correlations (Pearson's r, r^2), analysis of variance (η^2 , ω^2 , Cohen's f), and odds ratios (Odds Ratio, Risk Ratio). The interpretation of ES varies according to context, but general guidelines suggest that values such as 0.2, 0.5, and 0.8 in Cohen's d represent small, medium, and large effects, respectively. Finally, a practical example is included to illustrate the application of these measures and how SS and ES can lead to contrasting conclusions.

Keywords: Effect sizes; statistical significance; JASP.

Fecha de recepción: 06/11/2024

Fecha de aceptación: 30/01/2025

Correspondencia: Lindsay Michelle Vázquez, Universidad de Granada, España

Email: lindsay@ugr.es

¹ Se ha elegido JASP en detrimento de JAMOV (el habitualmente utilizado en las píldoras metodológicas), dado que el primer software incorpora una mayor gama de pruebas de tamaño del efecto.

Puntos clave

Lo que se sabe

- El uso de pruebas de significación estadística o contrastes de hipótesis no siempre es el procedimiento más adecuado para denotar efectos diferenciales entre grupos.

Lo que aporta este trabajo

- Nociones básicas para determinar posibles efectos diferenciales entre dos grupos (a través de una prueba de significación estadística y el cálculo de tamaños del efecto) a partir de un supuesto ficticio y mediante empleo del *software libre JASP*.

Escenario práctico

Se ha administrado una escala sobre el grado de ansiedad a 20 estudiantes universitarios antes del periodo de exámenes del primer cuatrimestre. Se toma en consideración sólo el último ítem de la misma o ítem criterio, cuyas posibles respuestas eran las siguientes: 1: nula, 2: baja; 3: moderada, 4: alta y 5: muy alta. Además, hay que tener en cuenta que los 10 primeros estudiantes no practican ningún tipo de estrategia de relajación, mientras los 10 últimos acuden a un curso de yoga organizado por su Universidad. Con estos precedentes se obtuvieron los siguientes resultados:

- a) No practican ninguna estrategia de relajación: 1,2,2,3,3,3,3,4,4,5.
- b) Acuden al curso de yoga organizado por su Universidad: 1,1,1,1,2,2,2,3,3,4.

El objetivo es determinar si se producen, o no, efectos diferenciales entre los dos grupos tomando como referencia los niveles de ansiedad. Se decide para el cumplimiento de este objetivo desarrollar pruebas de significación y tamaños del efecto y comparar los resultados.

Introducción

En la investigación científica, la significación estadística ha sido durante mucho tiempo la piedra angular para determinar la relevancia de los resultados obtenidos en estudios de diversa naturaleza metodológica. Tradicionalmente, la prueba de hipótesis y los valores- p han dominado el proceso de toma de decisiones, estableciendo un umbral arbitrario (generalmente 0.05) para decidir si un resultado es estadísticamente significativo, conduciendo al rechazo de la hipótesis nula cuando el valor es igual o menor que 0.05 ($p \leq 0.05$). Sin embargo, esta práctica ha enfrentado críticas crecientes por su incapacidad para reflejar la magnitud del efecto y la incertidumbre inherente en los datos, llevando a los/as investigadores/as a buscar alternativas más robustas y explicativas. Y he aquí que la magnitud del efecto o el tamaño del efecto puede proporcionar información complementaria a la significación estadística. También se podría destacar la estadística bayesiana, cuya importancia le confiere entidad suficiente para un tratamiento específico que podría ser abordado en una futura píldora formativa similar a la presente.

Principales Debilidades de la Significación Estadística

No son pocos los trabajos donde la significación estadística ha sido cuestionada por sus limitaciones. A este respecto, el trabajo de Fernández-Cano y Fernández-Guerrero (2009) parece uno de los más eruditos, completos y exhaustivos, por lo menos en el contexto de la investigación educativa. Sin entrar en el grado de profundización de dicha obra destacamos con Barriopedro (2015), Fernández-Cano y Fernández-Guerrero (2009), Kirk (2002), Onwuegbuzie y Levin (2003), Pascual Llobel et al. (2004) y Silva-Acayguer (2016), entre otros, las dos principales razones por las cuales deberíamos replantearnos el uso exclusivo de la significación estadística.

La primera razón es la supeditación de la significación estadística al tamaño muestral objeto de análisis. Esto se debe a que el valor p depende, tanto del tamaño del efecto, como del tamaño de la muestra.

Un efecto muy pequeño puede resultar significativo si el tamaño de la muestra es lo suficientemente grande y viceversa, es decir, no ser estadísticamente significativo, cuando sustancialmente es un efecto grande si el tamaño muestral es pequeño. En contraste, el tamaño del efecto ofrece una medida directa de la magnitud de una diferencia o relación, independiente del tamaño de la muestra, proporcionando una visión más clara de la importancia práctica de los resultados.

La segunda razón de calado es la arbitrariedad de la marca-frontera del valor p (0.05). A este respecto, Rosnow y Rosenthal (1989) llegaron a afirmar que “seguramente Dios ama al 0.06 casi tanto como al 0.05”. La decisión dicotómica derivada del establecimiento de un valor arbitrario concreto aumenta la presión de los investigadores por rebasar este umbral, lo cual a menudo resulta en la “p-hacking” o dragado de datos en castellano y otras prácticas cuestionables donde quienes investigan pueden manipular los datos o las condiciones del estudio para alcanzar significación estadística, por no hacer referencia a la discordancia a la que se puede llegar entre las vías de la significación estadística y significación sustantiva (Rodríguez-Sabiote et al, 2001).

Tamaños del Efecto como Alternativa a la Significación Estadística

Concepto

El concepto de tamaño del efecto tiene sus raíces en el desarrollo de la estadística y la psicometría en el siglo XX. Jacob Cohen, un psicólogo y estadístico, fue una figura clave en la popularización del tamaño del efecto, especialmente a través de su libro “Statistical Power Analysis for the Behavioral Sciences” cuya primera edición fue publicada en 1969. Antes de Cohen, el énfasis principal en la estadística aplicada se centraba en la significación estadística. Sin embargo, Cohen abogó por la importancia de evaluar la magnitud de los efectos observados, no solo si estos eran estadísticamente significativos.

En este contexto, el tamaño del efecto es una medida cuantitativa de la magnitud de un fenómeno. Indica la amplitud o intensidad de una relación entre variables, una diferencia entre grupos, o un cambio en una variable en el tiempo. A diferencia del valor p , que solo indica si un efecto es probable que exista, el tamaño del efecto proporciona una medida del impacto práctico, sustantivo o clínico. Manejamos estos términos diferentes (práctico, sustantivo o clínico) acerca de la medida del impacto de los tamaños del efecto, aunque en realidad se refieren a la presencia de impactos a tener en cuenta más allá de la significación estadística, pero dependiendo de la disciplina en que se implemente dicho efecto. Independientemente de la disciplina en que implementemos la herramienta del tamaño del efecto, debemos destacar una serie de rasgos distintivos que lo caracterizan, a saber:

- **Magnitud:** Proporciona una medida de cuán grande es un efecto, independientemente del tamaño muestral.
- **Por tanto, y como consecuencia de lo anterior, tenemos como rasgo distintivo la Independencia del Tamaño de la Muestra:** A diferencia del valor p , el tamaño del efecto no depende del tamaño de la muestra.
- **Comparabilidad:** Permite la comparación entre estudios y contextos diferentes, facilitando el meta-análisis, dado que un tamaño del efecto es una puntuación estandarizada z , al menos, en el caso de la comparación de medias.
- **Interpretabilidad:** Ofrece una interpretación más intuitiva y práctica de los resultados estadísticos.

Tipologías de Tamaños del Efecto

Existe un consenso más o menos extendido en la comunidad científica, que ha centrado su interés en la aplicación e interpretación de los tamaños del efecto, entre otros, Coe y Merino-Soto (2002), Cohen (2008) y Morales-Vallejo (2008), en afirmar que dichos tamaños pueden ser clasificados como se expone a continuación. No obstante, se advierte a quienes leen que no se trataría de una taxonomía única ni exhaustiva.

Basados en Diferencias de Medias

a) d de Cohen: mide la diferencia estandarizada entre dos medias usando la desviación estándar combinada de los grupos. Es útil para tamaños de muestra moderados a grandes.

b) g de Hedges: similar al d de Cohen, pero incluye una corrección para el sesgo en tamaños de muestra pequeños.

c) Δ de Glass: utiliza la desviación estándar del grupo control, adecuado cuando las desviaciones estándar de los grupos son significativamente diferentes y se considera que la desviación del grupo control refleja mejor la escala de medida (al no verse afectada por los efectos del tratamiento).

d) Diferencia de medias no estandarizada: presenta los resultados en la escala original, lo cual puede facilitar la interpretación. Una opción a tener en cuenta en casos en los que no se considera necesaria o aconsejable la estandarización.

Basados en Correlaciones

a) r de Pearson: mide la fuerza y dirección de la relación lineal entre dos variables cuantitativas. Rango interpretativo de -1 a +1. Existen otros índices adaptados a variables de naturaleza no cuantitativa, cuya interpretación es muy similar.

b) R^2 (Coeficiente de Determinación): proporción de la varianza en la variable dependiente que es explicada por la variable independiente.

Basados en Análisis de Varianza (ANOVA)

a) η^2 (Eta Cuadrática): mide la proporción de la varianza total atribuible a un factor. Indica el tamaño del efecto en términos de porcentaje de la varianza explicada.

b) ω^2 (Omega Cuadrática): similar a η^2 pero ajusta el sesgo en tamaños de muestra pequeños, proporcionando una estimación más realista del tamaño del efecto.

c) f de Cohen: utilizado en ANOVA para medir la magnitud de las diferencias entre medias ajustadas por la varianza dentro de los grupos.

Basados en Probabilidades o Cuotas

a) *Odds Ratio* (Razón de Cuotas): se calcula como la razón de la cuota de un evento en un grupo en comparación con otro grupo. Común en estudios de casos y controles, además de otros diseños.

b) *Risk Ratio* (Razón de Riesgos): cociente del riesgo (es decir, la frecuencia relativa o probabilidad) de un evento entre dos grupos. Utilizado en estudios de cohortes, entre otros diseños.

c) *Risk Difference* (Diferencia de Riesgos): consiste en restar el riesgo de un evento en los dos grupos comparados. El valor inverso de este índice se denomina Número Necesario de Tratamiento, y constituye un indicador útil del impacto de una intervención. Utilizado en diseños de cohortes, entre otros.

Interpretación de los Tamaños del Efecto

La interpretación del tamaño del efecto depende del contexto y la disciplina en que se esté implementando, pero existen guías generales que, en todo caso, pueden servir como punto de partida para una posible interpretación, aunque que no deben ser consideradas como axiomas. De esta forma, según Cohen (1992) se puede afirmar que el d de Cohen y otros índices similares (g de Hedges y Δ de Glass) pueden ser considerados como: Pequeño (0.2): Indica un efecto pequeño, pero significativo; Mediano (0.5): Representa un efecto moderado y Grande (0.8): Señala un efecto grande y fácilmente observable. No obstante, a lo explicitado, al final de este trabajo se propondrá una tabla de valores de tamaños de efecto con intervalos de valores más claramente etiquetados para su correcta interpretación.

Para r de Pearson (valor absoluto): Pequeño (0.1 a 0.25): Correlación débil; Moderado (0.25 a 0.40): Correlación moderada y Fuerte (>0.40): Correlación fuerte.

Finalmente, en relación con los tamaños del efecto basados en porcentajes de varianza explicada, suele considerarse que un valor en torno a 0.10 indica un efecto de magnitud baja-media, mientras que un valor en torno a 0.25 denota ya un efecto grande.

Desarrollo del ejemplo práctico propuesto con JASP

Cómo acceder a la matriz de datos creada previamente

Una vez elaborada la plantilla o matriz de datos basada en el escenario práctico contemplado, se abre el programa JASP (figura 1) y se clicla en la parte superior izquierda de las tres rayitas horizontales.

Figura 1

Página inicial del programa Jasp



Una vez allí, hay que buscar la ruta para llegar al archivo de interés (figura 2), que en este caso está elaborado en formato .sav de SPSS (ESPIRAL.sav), totalmente compatible con JASP, procediendo luego a abrirlo (figura 3).

Figura 2

Pantalla de apertura del archivo

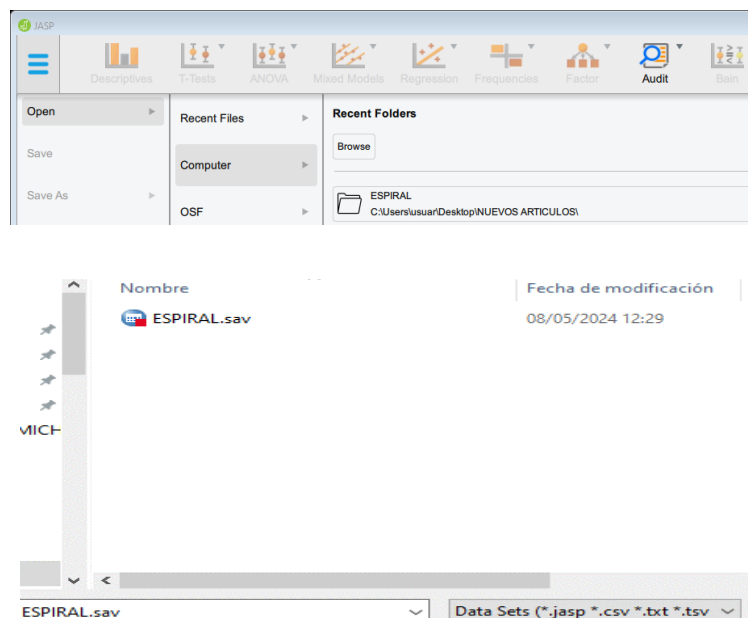
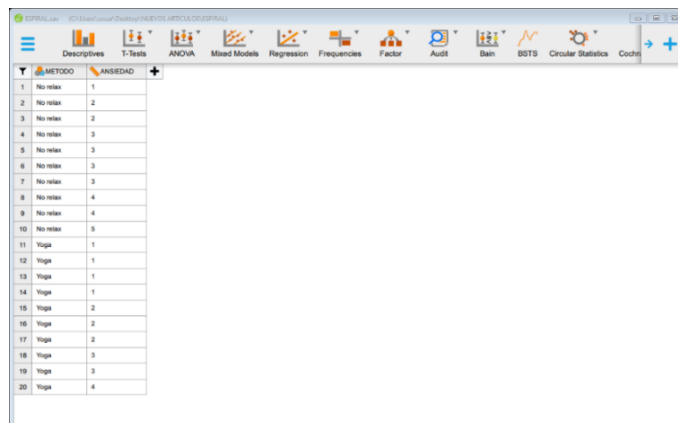


Figura 3

Pantalla de archivo abierto



Cómo acceder al procedimiento de análisis para responder al objetivo planteado en el escenario práctico

A modo introductorio, se van a calcular las medias aritméticas de cada grupo. Estos datos nos ayudarán a interpretar los resultados obtenidos. Para ello ir a *Descriptives* → *Descriptive statistics* → (figura 4) Colocar en la ventana de *Variables*: ANSIEDAD y en la ventana de *Split*: Método. Finalmente, sólo se marca *Mean* y se obtiene la tabla 1.

Figura 4

Pantalla de descriptive statistics

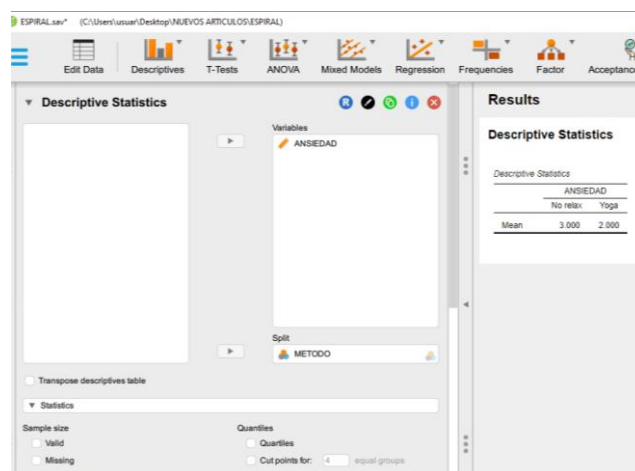


Tabla 1

Medias aritméticas de los dos grupos (no relax vs yoga) en la variable ansiedad

	Group	N	Mean
ANSIEDAD	No relax	10	3.000
	Yoga	10	2.000

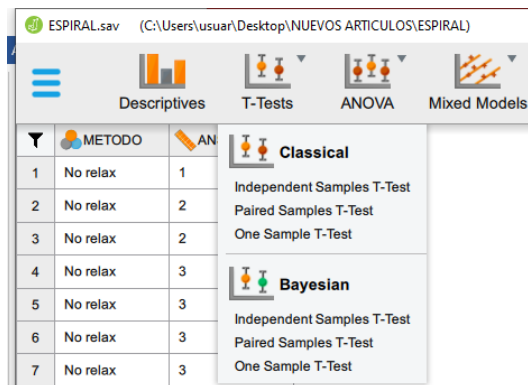
Note. *N* = tamaño de la muestra; *Mean* = media

Por otra parte, para el desarrollo de las pruebas de significación estadística, así como los tamaños del efecto, ir a (figura 5 y 6):

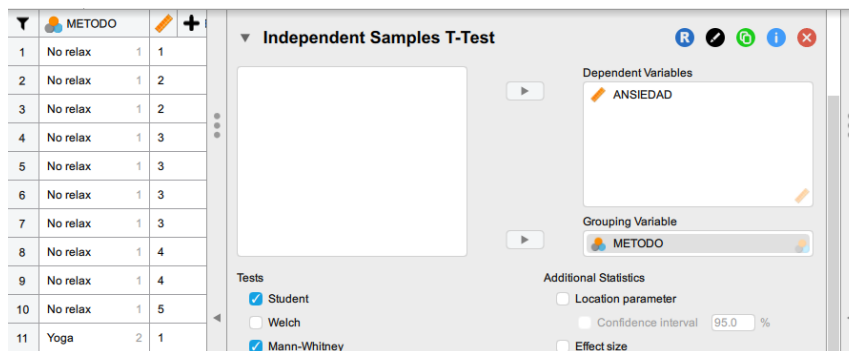
→*T-Test*→*Independent Sample t-test*

Figura 5

Acceso a pantalla de independent sample t-test


Figura 6

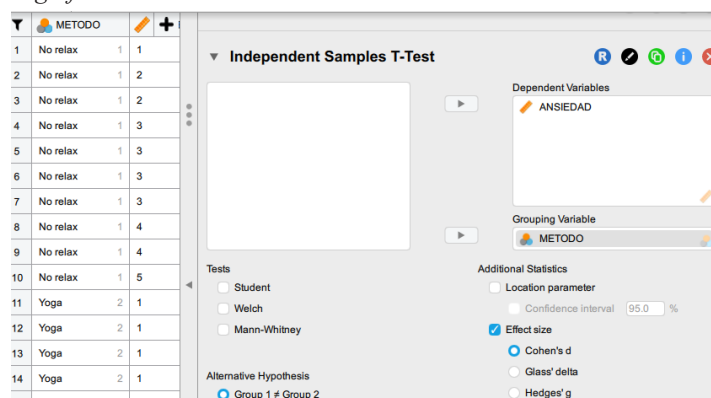
Pantalla de independent sample t-test



Tras esta acción habría que marcar las pruebas de significación estadística t de Student (de naturaleza paramétrica²) y su alternativa de carácter no paramétrica (U de Mann-Whitney). Adicionalmente, a modo de muestra inicial, aunque se desarrollará con más detalle más adelante, también se puede marcar *Effect size* con el estadístico *Cohen's d* (figura 7).

Figura 7

Pantalla de pruebas de significación estadística



² Pruebas donde se deben cumplir los supuestos de normalidad, independencia y, en su caso, homocedasticidad (o esfericidad, si se trata de una prueba F de ANOVA de medidas repetidas).

Tras estas acciones se mostrarán los resultados de la tabla 2.

Tabla 2

Resultados de las pruebas basadas en la significación estadística (t de Student y Mann-Whitney)

	Test	Statistic	df	P
ANSIEDAD	Student	2.023	18	0.058
	Mann-Whitney	74.000		0.067

Note. df = grados de libertad; p = valor de significación;

Si el interés se centrara en implementar medidas alternativas al tamaño d de Cohen se podrían marcar, adicionalmente, los estadísticos *Glass's delta* y *Hedges' g*. Dichos estadísticos difieren en el modo de calcular el denominador de la expresión anterior (d de Cohen), pero conceptualmente consisten también en una estandarización de la diferencia observada. Los resultados tras esta acción serían los de las tablas 3, 4 y 5.

Tabla 3

Descripción del contenido de la tabla

Independent Samples T-Test					
	T	df	p	Cohen's d	SE Cohen's d
ANSIEDAD	2.023	18	0.058	0.905	0.491

Note. Student's t-test; df = grados de libertad; p = valor de significación; SE = error estándar

Tabla 4

Resultados de la prueba basada en el tamaño del efecto de Δ de Glass

Independent Samples T-Test					
	t	df	p	Glass' delta	SE Glass' delta
ANSIEDAD	2.023	18	0.058	0.949	0.495

Note. Glass' delta uses the standard deviation of group Yoga of variable METODO.

Note. Student's t-test; df = grados de libertad; p = valor de significación; SE = error estándar

Tabla 5

Resultados de la prueba basada en el tamaño del efecto de "g" de Hedges

Independent Samples T-Test					
	t	df	p	Hedges' g	SE Hedges' g
ANSIEDAD	2.023	18	0.058	0.866	0.467

Note. Student's t-test. df = grados de libertad; p = valor de significación; SE = error estándar

Cómo Interpretar los Resultados Obtenidos

Para interpretar los resultados obtenidos se van a comentar cada una de las cuatro tablas explicitadas. De esta forma, en primer lugar, tomando en consideración las tablas 1 y 2 podemos apreciar como ambos estadísticos, es decir, t de Student = 2.02 y U de Mann-Whitney = 74. En este caso, para dichos valores con un nivel de confianza del 95% bilateral y para 18 grados de libertad (df , 20 participantes -2) se habrían obtenido probabilidades estadísticas $p > 0.05$, exactamente, $p = 0.058$ para la prueba t y $p = 0.064$ para la prueba U de Mann-Whitney.

Con estos precedentes se puede afirmar que no existen argumentos estadísticos para garantizar que las medias obtenidas (grupo yoga= 2 vs grupo no relax =3) sean estadísticamente diferentes, dado que en ambos casos las probabilidades estadísticas son > 0.05 . Hay que recordar a este respecto, que los valores de la variable dependiente van de 1 a 5 y que la diferencia entre los dos grupos, teniendo en cuenta el

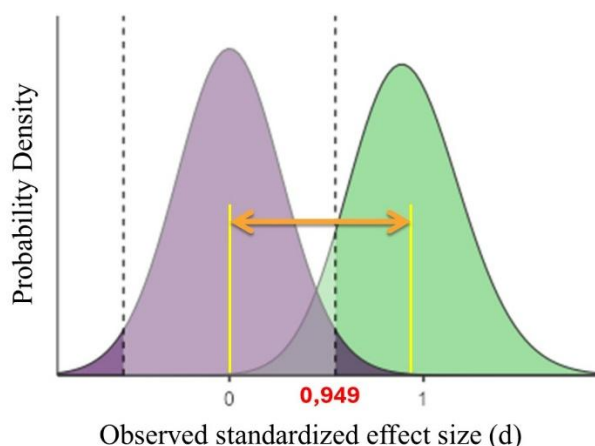
intervalo de dichas puntuaciones mínima vs máxima, es realmente importante, dado que es de 1 punto. Incluso así, la vía tradicional, es decir, la significación estadística, tanto desde una prueba paramétrica, como una no paramétrica han concluido que la diferencia de 1 punto no es lo suficientemente importante como para concluir que la formación de yoga haya sido eficaz en la disminución de la ansiedad.

Sin embargo, si se tienen en cuenta los resultados de los tamaños del efecto (tablas 3 a 5) las conclusiones pueden ser diametralmente opuestas. Los tres valores calculados están comprendidos de menor a mayor por 0.866 (g de Hegdes), 0.905 (d de Cohen y 0.949 (Δ de Glass). Ateniéndose a las etiquetas interpretativas de tales valores pueden considerarse como efectos grandes y fácilmente observables. Con estos resultados se puede concluir, al contrario que antes, que el grupo de yoga ha obtenido una media (2) en nivel de ansiedad sustantivamente menor que la del grupo de no relax (3).

Para que se entienda con mayor claridad, dado que un tamaño del efecto no es sino una puntuación estandarizada z , se puede afirmar que las medias del grupo no relax y las del grupo de yoga difieren entre sí por casi una $z=1$ o 1 desviación típica (si se toma como referencia, por ejemplo, la Δ de Glass=0.949). Si se quisiera representar gráficamente dicha distancia mediante dos curvas normales, representando cada una de ellas la distribución de cada grupo, se podría ver tal y como se representa en la figura 8.

Figura 8

Representación gráfica de la distancia estandarizada entre dos medias a través de distribuciones normales tomando el ejemplo implementado y el valor de la delta de Glass =0.949.



Es necesario aclarar que cuanto más solapadas estén dichas curvas, menor distancia hay entre los grupos que representan dichas curvas a nivel de diferencia estandarizada (menor tamaño del efecto). Por el contrario, cuanto más separadas estén mayor distancia hay entre los grupos que representan estas curvas a nivel de diferencia estandarizada (mayor tamaño del efecto).

Llegado este momento concluimos que los tamaños del efecto obtenidos son grandes; que implican una distancia estandarizada entre las medias de los grupos (yoga=2 y no relax=3) de casi una desviación típica³. Sin embargo, todavía no se logra alcanzar a ver cuán importante es esa distancia y por qué puede considerarse grande. Para ello se tendrá en cuenta la relación que existe entre los tamaños del efecto, en términos de percentiles, y el porcentaje de no solapamiento entre las distribuciones, tal y como se muestra en la tabla 6.

³ No confundir esta distancia con la diferencia entre las medias. Por otra parte, los tres tamaños del efecto se aproximan mucho a 1, pero no llegan a ese valor.

Tabla 6

Etiquetas interpretativas de los tamaños del efecto en términos de percentiles y porcentaje de no solapamiento entre las distribuciones

Etiqueta interpretativa	Tamaño del efecto	Percentil	% no solapado
Pequeño	0.0	50	0
	0.1	54	7.7
	0.2	58	14.7
	0.3	62	21.3
Medio	0.4	66	27.4
	0.5	69	33
	0.6	73	38.2
	0.7	76	43.0
Grande	0.8	79	47.4
	0.9	82	51.6
	1.1	84	55.4
	1.2	86	58.9
	1.3	88	62.2
	1.4	90	65.3
	1.5	91.9	68.1
	1.6	93.3	73.1
	1.7	94.5	75.4
	1.8	95.5	77.4
	1.9	96.4	79.4
	2	97.1	81.1
	2.5	99.0	87.7
	3	99.9	92.6

Fuente: Cohen (1992).

En este caso, imaginemos que se toma como valor de referencia el valor de la d de Cohen = 0.905 por su cercanía a 0.9. Con este resultado podríamos afirmar que el tamaño del efecto resultante es grande; que el 82% de los participantes del grupo no relax están por encima de la media en nivel de ansiedad del grupo de yoga y que, finalmente, la probabilidad porcentual (no solapamiento) de que un participante del grupo de yoga sea inferior en ansiedad al de no relax asciende al 51.6%.

Solución al Escenario Práctico

La lectura sosegada de este trabajo puede ofrecerte una visión que, probablemente, no conocías. Efectivamente, el conocimiento de procedimientos de análisis de datos en la comparación de grupos más allá de la significación estadística, estamos seguros de que te ha facilitado un nuevo escenario que puede complementar, cuando no sustituir a dicha aproximación en situaciones de controversia. La arbitrariedad en la que se asienta la significación estadística hace muy necesaria su complementación con otras herramientas, como los tamaños del efecto (abordados en el presente trabajo), así como los estadísticos bayesianos y la potencia de los contrastes, cuyo abordaje futuro, sin lugar a dudas, sería muy conveniente e interesante. En este contexto, recuerda que cuando tu interés científico se centre en denotar posibles efectos diferenciales entre grupos, como es el caso explicitado, debes implementar estadísticos clásicos, pero no olvides complementarlos con los tamaños del efecto. Nuestro consejo se basa en que puede que existan divergencias en las conclusiones derivadas de los análisis implementados mediante la vía significación estadística vs magnitud del efecto. En nuestro caso, al haberse reportado tales divergencias hemos tomado la decisión de quedarnos con la vía de la magnitud del efecto y concluir que el estudiantado que ha cursado la estrategia de yoga realmente exhibe niveles de ansiedad sustantivamente menores y, por tanto, a tener en cuenta.

Conclusiones

El uso de pruebas de contraste de hipótesis o pruebas de significación, tanto de naturaleza paramétrica como no paramétrica, es la primera opción de referencia cuando se pretende comprobar si existen efectos diferenciales entre grupos. Siendo una opción viable, legítima y estadísticamente fundamentada, no obstante, debe ser tomada con cautela, sobre todo cuando, como es el caso mostrado en este trabajo, las comparaciones se implementan con tamaños muestrales pequeños (efectos grandes no estadísticamente significativos), aunque también este sesgo se puede producir con tamaños muestrales grandes (efectos pequeños sí estadísticamente significativos).

La divergencia de conclusiones a las que puede llegarse por una vía (contrastes clásicos) y por otra (magnitud del efecto) implica que todo investigador/a que lleve a cabo comparaciones entre grupos debería cuestionarse los resultados obtenidos a partir de la significación estadística y comprobar si por la vía de la magnitud del efecto puede verificar la presencia de efectos diferenciales sustantivos.

Esto es lo que ha acontecido en el caso mostrado, o sea, la presencia de un efecto grande no estadísticamente significativo por efecto de una muestra pequeña. Esta conclusión ha sido cuestionada por los tamaños del efecto resultantes que han demostrado la presencia de un efecto sustantivo a tener en cuenta. En esta situación de divergencia la persona investigadora debe ser libre para elegir una u otra vía, pero honestamente y habida cuenta de las limitaciones de la significación estadística, es más razonable elegir la vía de la magnitud del efecto o también otra alternativa, igualmente razonable, como es la estadística bayesiana. Esta última será abordada en otra ocasión.

Contribución de cada Autor: Conceptualización, C.R.S., J.A.L.L., y M.S.M; escritura del manuscrito, C.R.S., L.M.V., y J.A.L.L.; escritura, revisión y edición, C.R.S., L.M.V., J.A.L.L., y M.S.M; supervisión, C.R.S., L.M.V., J.A.L.L., y M.S.M.

Financiación: Esta investigación ha recibido financiación del Ministerio de Ciencia, Innovación y Universidades. Formación de Profesorado Universitario (FPU22/01938).

Conflicto de Intereses: Las/os autoras/es declaran que no tienen conflicto de intereses.

Referencias

- Barriopedro, M. (2015). La significación estadística no es suficiente. *RICYDE. Revista Internacional de Ciencias del Deporte*, 11, (40) 101-103. DOI: <http://dx.doi.org/10.5232/ricyde2015.040ed>
- Coe, R. y Merino Soto, C. (2003). Magnitud del Efecto: Una guía para investigadores y usuarios. *Revista De Psicología*, 21(1), 145-177. <https://doi.org/10.18800/psico.200301.006>
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cohen, B.H. (2008). *Explaining psychological statistics*. John Wiley & Sons.
- Fernández-Cano, A. y Fernández-Guerrero, I. (2009). *Crítica y alternativas a la significación estadística en el contraste de hipótesis*. La Muralla.
- JASP Team (2024). *JASP (Version 0.18.3)* [Computer software].
- Kirk, R.E. (2001). Promoting Good Statistical Practices: Some Suggestions. *Educational and Psychological Measurement*, 61 (2), 213-218.
- Morales Vallejo, P. (2008). *Estadística Aplicada a las Ciencias Sociales*. Universidad Pontificia Comillas.
- Onwuegbuzie, A.J. y Levin, J. R. (2003). Without supporting statistical evidence, where would reported measures of substantive importance lead? To no good effect. *Journal of Modern Applied Statistical Methods*, 2, (1), 133-151.
- Pascual Llobell, J., Frías Navarro, M.D. y García Pérez, F. (2004). Usos y abusos de la significación estadística: propuestas de futuro (¿Necesidad de nuevas normativas editoriales?). *Metodología de las Ciencias del Comportamiento, Volumen Especial*, 465-469.

- Rodríguez-Sabiote, C., Lorenzo-Quiles, O. y Navarro-Hernández, J. J. (2001). Un ejemplo práctico sobre la discordancia entre la significación estadística y la significación sustantiva con relación a la decisión de la hipótesis nula. *Publicaciones*, 31, 173–186.
- Rosnow, R. L. y Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276-1284.
- Silva-Acayguer, L. (2016). Las pruebas de significación estadística. Seis décadas de fuegos artificiales. *Revista Nacional de Salud Pública*, 34 (3), 372-379.

Anexo

El tamaño del efecto como alternativa a las pruebas de significación estadística. Un ejemplo práctico con JASP

