

Impacto de la reforma estadística en los docentes de Psicología y de Educación: tamaño del efecto y de la muestra

L. Badenes-Ribera*; M. Pascual-Soler**; R. Gomez-Frias*, H. Monterde-i-Bor* & D. Frias-Navarro*

**Universitat de València (España)*

***ESIC Business & Marketing School (España)*

RESUMEN (ABSTRACT)

La estadística es una materia difícil de enseñar y aprender y hay evidencia que señala que su aplicación es deficiente en Psicología y Educación. Los errores metodológicos incluyen los aspectos de diseño, análisis, generación de informes e interpretación. Nuestra investigación tiene dos objetivos: evaluar si los profesores universitarios españoles estiman la potencia estadística en la fase de planificación de sus estudios y, por lo tanto, cómo definen el tamaño óptimo de la muestra. Y analizar en qué grado ha impactado la reforma estadística entre los profesores españoles de Psicología y Educación. La muestra está formada por 744 profesores de las disciplinas de Psicología y Educación, con una antigüedad media de 12.85 años (SD=9.23) en la universidad, procedentes de universidades públicas (84.9%) y privadas (13.2%). En consonancia con otros estudios, los principales resultados señalan la presencia de errores de interpretación de los valores p, la falta de planificación del tamaño de la muestra así como del tamaño del efecto. Por lo tanto, el impacto de las indicaciones de la reforma estadística en España aún no se ha producido, manteniéndose un modelo tradicional de la conducta metodológica del investigador. Investigación financiada por el 'Ministerio de Economía y Competitividad', España (EDU2011-22862).

Palabras clave: reforma estadística, cognición estadística, validez de conclusión estadística.

1. INTRODUCCIÓN

1.1 Problema/cuestión.

Nuestros resultados avalan lo que muchos investigadores señalan: la comprensión y la aplicación correcta de muchos conceptos estadísticos continúan siendo problemática y la aplicación de muchas técnicas estadísticas imprecisa y, además, sin atender a los requisitos o supuestos de aplicación de las pruebas que conlleva en muchos casos la invalidez de los resultados o conclusiones obtenidas. Confundir el nivel de significación de alfa con la probabilidad de que la hipótesis nula sea cierta, interpretar un resultado estadísticamente significativo como un resultado importante o útil y creer que el nivel de significación se mantiene de forma estática independientemente del número de comparaciones o pruebas que se ejecuten son cuestiones que hemos estudiado y debatido en nuestros trabajos y publicaciones.

El presente trabajo analiza si los profesores universitarios españoles consideran dos cuestiones básicas relacionadas con la validez de conclusión estadística: la potencia estadística y la planificación del tamaño óptimo de la muestra.

1.2 Revisión de la literatura.

La educación, la investigación y el conocimiento son tres retos fundamentales para el crecimiento de la sociedad del conocimiento donde también es necesaria la difusión de los resultados de investigación, de innovación, de cultura científica y de la educación superior universitaria a la propia sociedad

La calidad del conocimiento científico generado en una disciplina requiere que los investigadores planifiquen adecuadamente su investigación, la ejecuten eficientemente, analicen los datos correctamente, interpreten bien los resultados y presenten de forma clara y transparente las conclusiones. Sin embargo, los usos y abusos que los investigadores realizan del diseño estadístico y los problemas de validez de los hallazgos son cuestiones que rodean al proceso de investigación y que dificultan la acumulación del conocimiento basado en la mejor evidencia.

Como consecuencia de los usos y abusos de las pruebas estadísticas basadas en el contraste de la hipótesis nula entendida como efectos cero surgen las recomendaciones de los organismos como American Psychological Association (APA) y American Educational Research Association (AERA). El movimiento de la Medicina Basada en la Evidencia (MBE)

y, en general, el de la Práctica Basada en la Evidencia (PBE) tiene un punto de arranque en la toma de conciencia de los problemas vinculados al procedimiento de significación de la hipótesis nula como único medio para producir descubrimientos, enfatizando el uso de las revisiones sistemáticas y la estimación del tamaño del efecto.

La denominada reforma estadística cambia el punto de mira desde “cómo es de probable o improbable el resultado muestral” hacia dos cuestiones principalmente: “cómo es de grande el tamaño del efecto detectado” y “si se puede replicar”. Es decir, hay que ‘evaluar’ el valor del tamaño del efecto estimado y su utilidad (su grado de importancia práctica) y para ello es necesario considerar el contexto de la investigación y comparar los resultados de forma explícita y directa con los obtenidos en el área de investigación donde se enmarca el trabajo (pensamiento meta-analítico). Por pensamiento meta-analítico entendemos al igual que Thompson (2002) como aquel que a) facilita la formulación prospectiva de nuevas hipótesis y su diseño invocando explícitamente los tamaños del efecto previamente descubiertos y b) facilita la interpretación retrospectiva de los nuevos resultados en comparación directa con los tamaños del efecto previamente descubiertos por la investigación anterior. En una palabra, se facilita el proceso de integración y replicación, tan necesarios a la Ciencia, y que no han sido favorecidos por la aplicación rutinaria y excesivamente mecánica del contraste de hipótesis.

La última edición del Manual de la American Psychological Association (2010) mantiene y refuerza su énfasis en la denominada reforma estadística destacando el uso de los tamaños del efecto y sus intervalos de confianza y las técnicas bayesianas tratando de minimizar la confianza excesiva que los investigadores tienen sobre las pruebas de significación estadística y las decisiones dicotómicas apoyadas en los valores p de probabilidad.

Generalmente en el mundo científico se trabaja con muestras aleatorias y se necesita de la inferencia estadística para obtener conclusiones sobre la población y por ello la valoración de la potencia estadística a priori en la fase de planificación del estudio es fundamental. En la fase de planificación de una investigación es primordial plantearse qué tamaño muestral es necesario para garantizar de forma razonable el equilibrio entre la probabilidad de rechazar la hipótesis nula (siendo realmente verdadera, error de Tipo I) y el riesgo de no rechazarla (siendo realmente falsa, error de Tipo II). Es necesario, por lo tanto, planificar los riesgos de error de Tipo I (seleccionar alfa), de error de Tipo II (nivel aceptable de beta) y el tamaño del efecto poblacional esperado (estimado). Con esas tres cantidades ya

puede el investigador decidir a priori qué tamaño muestral (N) deberá tener el estudio para garantizar la validez de conclusión estadística. En otras palabras, los estudios de la potencia estadística a priori (validez de conclusión estadística) tratan de dar respuesta a la cuestión sobre el tamaño de la muestra: ¿cuántos sujetos deben formar la muestra?

En la práctica, el investigador intentará determinar el tamaño óptimo de la muestra para asegurar la representatividad, pero deberá considerar también la disponibilidad de recursos de diversa índole para la realización del trabajo de campo. Se trata de estimar el tamaño de muestra que sea óptimo para los objetivos del estudio. Si se estudian más elementos de los necesarios se están derrochando recursos y si se estudian menos elementos de los necesarios entonces el diseño estadístico no tendrá la suficiente potencia estadística o seguridad sobre los hallazgos y podría suceder que no se detecten efectos que realmente sí que existen en la población.

Por lo tanto, cuando se lleva a cabo un estudio es necesario que el investigador planifique adecuadamente la potencia estadística de las pruebas estadísticas que va a utilizar. Se trata de proteger a la validez de conclusión estadística. Para ello el investigador debe reflexionar sobre una serie de cuestiones metodológicas como:

- ¿Cuánta potencia estadística se desea?
- ¿Qué valor probable del tamaño del efecto (en la población) es el que trata de detectar en su estudio? O ¿qué valor de tamaño del efecto se considera trivial?
- ¿Qué criterio de nivel de significación se va a utilizar en el estudio (alfa)?
- ¿Qué prueba estadística se va a emplear?

Por convención se considera que el valor máximo de alfa es .05 y el valor de la potencia estadística deseada es al menos de .80 (en el 80% de las ocasiones se rechazará la hipótesis nula siendo realmente falsa). El valor del tamaño del efecto puede ser estimado a partir de un estudio de meta-análisis sobre el tema objeto de análisis o mediante un estudio piloto. Si no se dispone de un trabajo de meta-análisis o del estudio piloto también se puede estimar la potencia con los valores tradicionales de Cohen (1988) de tamaño del efecto pequeño ($d = 0.2$), mediano ($d = 0.5$) y grande ($d = 0.8$) cuando se trata de diferencia de medias. Si se dispone de un tamaño del efecto medio de un trabajo de meta-análisis siempre será la mejor opción ya que se contextualiza el tamaño del efecto en un área concreta de investigación.

En definitiva, planificar un valor de 0.8 de potencia estadística está unido a la planificación del tamaño de la muestra, previsto un valor del tamaño del efecto y fijado un nivel de alfa a priori.

1.3 Propósito.

Los objetivos de nuestro trabajo son dos: 1) evaluar si los profesores universitarios españoles estiman la potencia estadística en la fase de planificación de sus estudios y, por lo tanto, cómo definen el tamaño óptimo de la muestra y 2) analizar en qué grado ha impactado la reforma estadística entre los profesores españoles de Psicología y Educación.

2. METODOLOGÍA

2.1. Descripción del contexto y de los participantes

La muestra está formada por 744 profesores de las disciplinas de Psicología (n=472, 63.4%) y Educación (n=272, 36.6%), con una antigüedad media de 12.85 años (SD=9.23) en la universidad, procedentes de universidades públicas (84.9%) y privadas (13.2%). Los hombres representan el 48.9% y las mujeres el 51.1%. El proceso de selección de la muestra fue equivalente al adoptado por otros estudios que han trabajado con poblaciones de personal docente-investigador (Gordon, 2001; Mittag & Thompson, 2000; Monterde-i-Bort, Frias-Navarro & Pascual-Llobell, 2010).

2.2. Instrumento de encuesta

El instrumento de encuesta recoge información relacionada con las variables de sexo y edad, facultad de psicología/Facultad de Educación, antigüedad en la universidad y tipo de universidad como pública o privada.

La encuesta tiene un apartado dedicado a la conducta del investigador vinculada con la valoración de la potencia estadística a priori y la planificación del tamaño de la muestra. Las cuestiones son las siguientes:

- a) Potencia estadística y tamaño de la muestra: cuando planifica un estudio ¿estima a priori el tamaño de la muestra?, ¿qué tipo de estrategia utiliza cuando desea planificar el tamaño de la muestra de un estudio? Y en su opinión ¿qué función tiene estimar la potencia estadística a priori?
- b) Impacto de la reforma estadística entre los profesores españoles de Psicología y Educación: podría señalar el nombre de algunos estadísticos del tamaño del efecto, cuando ejecuta una prueba estadística considera prioritario informar siempre de la

significación estadística y podría señalar qué términos conoce adecuadamente de la siguiente lista (desviación típica, gráfico de sedimentación, forest plot, ANOVA, funnel plot, ejes rotatorios (Displax), correlación, meta-análisis, análisis de regresión y tamaño del efecto).

2.3. Procedimiento

Se realizó una base de datos con todo el personal docente e investigador de las universidades públicas y privadas españolas cuyos datos estaban recogidos en las páginas web de las facultades de Psicología y Educación. Posteriormente se seleccionó una muestra aleatoria de sujetos (margen de error del 5%) y se les envió la encuesta a través de Internet.

3. RESULTADOS

Potencia estadística y tamaño de la muestra:

El estudio de la planificación de la potencia estadística se realizó con tres cuestiones cuyos porcentajes de respuesta se detallan en la Tabla 1.

Tabla 1. Porcentajes de respuesta respecto a la planificación de la potencia estadística

Cuestión	No	Sí		
<i>Quando planifica un estudio ¿estima a priori el tamaño de la muestra?</i>	15.1% (112)	66.9% (498)		
	Trata de conseguir el mayor número posible de participantes	Utiliza un programa informático o tablas que permiten estimar el tamaño de la muestra en función de ciertos criterios estadísticos	Procura que la muestra represente las características de la población	No utiliza ningún tipo de estrategia porque no forma parte de sus intereses de investigación
<i>¿Qué tipo de estrategia utiliza cuando desea planificar el tamaño de la muestra de un estudio?</i>	21.2% (158)	25.9% (193)	39.4% (293)	4.6% (34)
	Ajustar el valor del nivel de significación o alfa	Explorar la fiabilidad de las escalas	Estimar el tamaño de la muestra	
<i>¿Qué función tiene estimar la potencia estadística a priori?</i>	34.1% (254)	11.2% (83)	41.5% (309)	

Impacto de la reforma estadística entre los profesores españoles de Psicología y Educación:

El estudio del impacto de la reforma estadística se realizó con tres preguntas cuyos porcentajes se detallan en la Tabla 2.

Tabla 2. Porcentajes de respuesta respecto al impacto de la reforma estadística

Cuestión	No conoce	Sí	
<i>Podría señalar el nombre de algunos estadísticos del tamaño del efecto</i>	41.9% (312)	58.1% (432)	
	No	Sí y utiliza expresiones como $p < 0.05$, $p > 0.05$	Sí y utiliza expresiones con el valor p de probabilidad exacto
<i>Cuando ejecuta una prueba estadística considera prioritario informar siempre de la significación estadística</i>	7.8% (58)	52.6% (391)	39.7% (295)
<i>Podría señalar qué términos conoce adecuadamente de la siguiente lista:</i>			
• Correlación	96.5% (718)		
• Desviación típica	96.4% (717)		
• ANOVA	89.9% (669)		
• Análisis de regresión	85.3% (635)		
• Meta-análisis	76.9% (572)		
• Tamaño del efecto	71.4% (531)		
• Gráfico de sedimentación	38.7% (288)		
• Ejes rotatorios (Displax)	12.5% (93)		
• Forest Plot	9.7% (72)		
• Funnel Plot	5.5% (41)		

4. CONCLUSIONES

En las Ciencias Sociales como la Psicología y la Educación, las técnicas de inferencia estadística se utilizan para dar respuesta a una serie de hipótesis a través del análisis de los datos recogidos en el estudio. Cuando se diseña un estudio hay un conjunto de decisiones que son críticas y están basadas en información incompleta o incierta antes de recoger los datos como es el valor del tamaño del efecto que se obtendrá en el estudio. Una de estas decisiones críticas es planificar a priori el tamaño de muestra necesario para alcanzar los objetivos de la investigación, es decir, valorar el valor de la potencia estadística.

La comprobación de hipótesis estadísticas exige el conocimiento de los fundamentos de la herramienta de la estadística y del proceso del diseño de investigación. En concreto, los elementos vinculados con la validez de conclusión estadística son fundamentales. Planificar el tamaño de la muestra (dado un valor de potencia estadística, un tamaño del efecto y un nivel

de significación) supone valorar el esperable tamaño del efecto y desarrollar el pensamiento meta-analítico (Frias-Navarro, 2011). El grupo de trabajo de inferencia estadística de la American Psychological Association ya enfatizaba la necesidad de medir el tamaño del efecto en los estudios primarios y el mismo manual de la American Psychological Association (2010) demanda la estimación del tamaño del efecto junto a los valores p de probabilidad de las pruebas estadísticas, realizando una interpretación contextualizada. Los valores p son ampliamente utilizados en Psicología y Educación pero es cierto que muchas veces de forma inadecuada, realizando interpretaciones erróneas que no aportan los análisis (Monterde-i-Bort, Frias-Navarro & Pascual-Llobell, 2010; Monterde-i-Bort, Pascual-Llobell & Frias-Navarro, 2006).

Los resultados de nuestro estudio señalan que el 67% aproximadamente de los docentes e investigadores españoles de las facultades de Psicología y Educación dicen estimar a priori el tamaño de la muestra. Sin embargo, la mayoría (49.4%) lo hacen con el propósito de obtener una muestra representativa y no para disminuir el error de Tipo II o error beta. De este modo, se confunde la cuestión de la representatividad de la muestra con la potencia estadística o probabilidad de rechazar la hipótesis nula siendo realmente falsa.

Respecto al impacto de la reforma estadística, los resultados señalan que aproximadamente el 42% de los participantes desconoce el nombre de un estadístico del tamaño del efecto. Sin embargo, el 77% dice conocer adecuadamente lo que significa el tamaño del efecto. Es decir, se conoce el término pero no se ha incluido en la conducta del investigador dado que no se aplica pues se desconoce qué tipo de estadístico puede informar del valor del tamaño del efecto. Y cuando ejecutan una prueba estadística, el 53% de los participantes prefieren señalar los resultados del valor p de probabilidad como mayor o menor al alfa. Las recomendaciones del manual de la American Psychological Association (2010) son claras: siempre aportar el valor exacto de probabilidad vinculado a la prueba estadística para evitar confusiones e interpretaciones inadecuadas.

Los cambios que desde ya hace años se postulan como básicos en la práctica estadística no llegan a consolidarse entre la conducta de los investigadores, en los manuales de metodología, en las políticas editoriales de las revistas ni tampoco son incorporados en la formación continua de los profesionales. Por todo ello es necesaria la elaboración de nuevos materiales de aprendizaje y de investigación, basados en los resultados de estudios sobre razonamiento estadístico, y, por supuesto, la difusión de los contenidos e implicaciones de la

reforma estadística entre los profesionales, docentes y alumnos. Las estrategias para planificar el tamaño de la muestra puede realizarse con las tablas de Cohen (1988) o puede determinarse con programas estadísticos específicos como el G*Power (Erdfelder, Faul, & Buchner, 1996). Para determinar el tamaño del efecto poblacional y así poder planificar el tamaño de la muestra, se puede optar por tres aproximaciones: 1) utilizar el valor del tamaño del efecto que se considera importante para detectar, 2) utilizar los valores convencionales de Jacob Cohen (1988) y 3) ejecutar una investigación piloto que determine el posible valor del tamaño del efecto, o si es posible trabajar con el tamaño del efecto medio de un estudio de meta-análisis (Gravetter & Wallnau, 2006; Howell, 2007).

El siglo XXI se inicia después de una profunda reflexión sobre la significación estadística, el tamaño del efecto y la significación práctica, la importancia de la potencia estadística en el procedimiento de significación estadística, el desarrollo de la técnica de meta-análisis como una estrategia integradora, las limitaciones sobre las pruebas de significación estadística y el planteamiento de contrastes específicos y de efectos mínimos. Ya no es suficiente conocer si hay alguna relación entre las variables, ahora es esencial saber la magnitud de la relación o si es suficientemente grande y contextualizar el valor de dicha magnitud en un área concreta de investigación. El cambio propuesto no es sencillo.

Este trabajo se enmarca dentro de la línea de investigación sobre educación estadística y cognición que nuestro equipo de investigación viene desarrollando desde hace años en el Departamento de Metodología de las Ciencias del Comportamiento de la Universitat de València. En concreto dentro del proyecto de investigación “*Impacto de la reforma estadística en Educación y Psicología: de la significación estadística a la estimación de efectos*” (Frias-Navarro, D., Monterde-i-Bort, H, Pascual Soler, M. y Pascual Llobell, J.). Proyecto subvencionado por el Ministerio de Ciencia e Innovación español, Dirección General de Investigación y Gestión del Plan Nacional de I+D+i (EDU2011-22862).

5. REFERENCIAS BIBLIOGRÁFICAS

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum Associates.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Finch, S., Thomason, N., & Cumming, G. (2002). Past and future American Psychological Association guidelines for statistical practice. *Theory and Psychology*, 12, 825-853.
- Frias-Navarro, D. (2011). Reforma estadística: tamaño del efecto. En D. Frias-Navarro, *Técnica estadística y diseño de investigación*. Valencia: Ediciones Palmero.
- Gordon, H. R. D. (2001). American Vocational Education Research Association members' perceptions of statistical significance tests and other statistical controversies. *Journal of Vocational Educational Research*, 26, 1-18.
- Gravetter, F.J., & Wallnau, L.B. (2006). *Statistics for the behavioral sciences* (7th ed). Belmont, CA: Wadsworth Publishing.
- Howell, D. C. (2007). *Statistical methods for psychology* (6th ed). USA: Thompson/Wadsworth.
- Mittag, K. C., y Thompson, B. (2000). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. *Educational Researcher*, 29, 14-20.
- Monterde i Bort, H., Pascual, J., & Frias-Navarro, D. (2006). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones. *Psicothema*, 18, 848-856.
- Monterde-i-Bort, H., Frias-Navarro, D., & Pascual, 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.
- Nickerson, R. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Chichester, UK: Wiley.

Thompson, B. (2002). What future quantitative social science research could look like:
Confidence intervals for effect size. *Educational Researcher*, 31, 25-32.