Metodología de las Ciencias del Comportamiento
2002. Volumen Especial

ISSN 1575-9105

La hipótesis nula y la significación práctica

Frías , M. D.; Pascual, J. y García, J.F.
Universidad de Valencia

Durante la década de los noventa del siglo XX se han publicado debates y opiniones (se puede consultar una revisión exhaustiva en Nickerson, 2000) que han cuestionado, en ocasiones duramente, la aplicación del procedimiento de significación de la hipótesis nula (Null Hypothesis Significance Testing, NHST) como instrumento para el progreso del conocimiento científico. Las críticas al procedimiento incluyen desde las falsas concepciones sobre la información que facilita el procedimiento, provocando interpretaciones erróneas de sus resultados, hasta su uso inadecuado como medio de obtención de datos de significación práctica, sustantiva o clínica. El objetivo de nuestro estudio es abordar la problemática no resuelta aún sobre la utilización apropiada de NHST dentro del área aplicada de la comprobación de efectos del tratamiento y del cambio clínico así como analizar la dificultad de la especificación a priori del valor del tamaño del efecto y analizar las diferentes propuestas que actualmente se proponen al procedimiento. Entre dichas propuestas destaca la estimación del tamaño del efecto y el planteamiento de tamaños de efecto mínimo y datos normativos que enriquecen el diseño de investigación tanto en relación a su validez como en su aportación al progreso científico de la disciplina objeto de estudio. Los resultados de los trabajos de meta-análisis son especialmente útiles para la planificación de los valores del tamaño del efecto y para los análisis a priori de potencia de la prueba estadística.

During the decade of ninety in the XX century debates and opinions have been published (it can consult himself an exhaustive revision in Nickerson, 2000) that have questioned, in occasions difficultly, the application of the procedure of significance of the null hypothesis (Null Hypothesis Significance Testing, NHST) like instrument for the progress of the scientific knowledge. The critics to the procedure include from the false conceptions about the information that facilitates the procedure, causing erroneous interpretations of their results, until their inadequate use as half of obtaining of data of practical significance, sustantiva or clinic. The objective of our study is to not approach the problem still solved on the appropriate use of NHST inside the applied area of the confirmation of effects of the treatment and of the clinical change as well as to analyze the difficulty of the specification a priori of the value of the size of the effect and to analyze the different proposals that at the moment they intend to the procedure. Among this proposals the estimate of the size of the effect and the position of sizes of minimum effect and normative data that enrich the so much investigation design in relation to their validity highlights like in their contribution to the scientific progress of the discipline study object. The results of the goal-analysis works are specially useful for the planning of the values of the size of the effect and for the analyses a priori of power of the statistical test.

Dentro del área de la psicología aplicada interesa comprobar los efectos de los tratamientos donde lo importante no es únicamente mostrar que los tratamientos difieren o lo que es lo mismo si el tratamiento tuvo *algún* efecto sino que el interés se encuentra en señalar los beneficios de un tratamiento sobre otro, es decir, se desea conocer si el tratamiento tiene el efecto que el investigador plantea en su hipótesis científica (Fowler, 1985), apoyando la significación práctica o importante del cambio siguiendo el juicio del investigador. En general, en los estudios se plantea que el impacto del tratamiento producirá mayores cambios en la respuesta de los sujetos respecto al grupo de control no sometido a la intervención. El impacto del tratamiento es el denominado tamaño del efecto. Cuando se trata de analizar el denominado cambio clínico el efecto formulado en la hipótesis científica recoge el cambio hacia los valores de normalidad dentro del área psicológica sometida a intervención. En estos casos puede suceder que el cambio estadísticamente significativo no indique el verdadero valor terapéutico, prevaleciendo la importancia de la significación sustantiva, práctica o la denominada significación clínica entendida como la magnitud del cambio atribuida al tratamiento terapéutico (Kendall, Flannery-Schroeder y Ford, 1999) que permite que el funcionamiento del sujeto pueda ser considerado normal. Ya sea como estimación de ciertos efectos o como cambio clínico, estamos hablando de analizar determinados efectos cuya cuantía está determinada por el modelo teórico y el campo psicológico en el que se trabaje.

De acuerdo con estos planteamientos de investigación, el procedimiento estadístico clásico basado en la hipótesis de nulidad de efectos con valor cero (hipótesis *nil* o vacía en términos de Cohen, 1994) no puede dar respuesta a las necesidades planteadas por hipótesis de nulidad con efectos distintos de cero (*non-nil* hipótesis) como pueden plantearse por ejemplo dentro del área de la valoración de programas o en el contexto del cambio clínico. Cuando se planifican diseños de investigación que están especialmente interesados en la comprobación de efectos del tratamiento, el cálculo del tamaño del efecto ayuda a la comprensión sustantiva de los resultados, complementando (sustituyendo, según algunos autores) el procedimiento tradicional de la comprobación de la significación estadística. No sólo por la necesidad de valorar el efecto del tratamiento desde el punto de vista del análisis del cambio práctico o clínico sino también debido a la interpretación incorrecta y uso inadecuado de las pruebas de significación estadística, la integración del análisis sustantivo de las diferencias detectadas entre los grupos debe ser incorporada al diseño de investigación. Analicemos brevemente los tres problemas más destacados que están vinculados con el uso e interpretación de los procedimientos clásicos de significación estadística.

Correspondencia: María Dolores Frías Navarro
Departamento de Metodología de las CC. del Comportamiento. Facultad de Psicología.
Universidad de Valencia
Avda. Blasco Ibáñez 21. 46010 Valencia.
E-mail: friasnav@uv.es

¿Qué significa utilizar el procedimiento de significación de la hipótesis nula? Implica poner en marcha el proceso de rechazar o mantener una hipótesis nula concreta basada en una serie de consideraciones teóricas planteadas a priori y en los valores *p* de probabilidad vinculados a la prueba estadística empleada con el objetivo de obtener conclusiones respecto a una hipótesis alternativa. Y desde este mismo planteamiento surge una de las críticas más difíciles de responder:

el procedimiento no especifica realizar predicciones sobre la hipótesis de investigación o científica sino que las predicciones se realizan sobre la hipótesis estadística de nulidad y será su rechazo la que dará crédito a la hipótesis de investigación que ni es especificada ni es comprobada con dicho procedimiento estadístico.

1. Interpretación incorrecta del valor de probabilidad *p.*

Se interpretan incorrectamente las probabilidades condicionales de resultados empíricos, $P(D/H_0)$, asociado al valor *p* de probabilidad, como probabilidades condicionales de hipótesis, $P(D/H_0)$ (Cohen, 1994), buscando respuestas que la técnica no puede ofrecer.

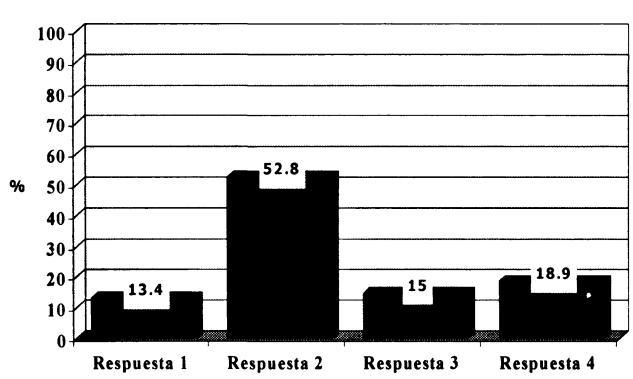
2. El valor *p* no informa del efecto del tratamiento.

La técnica no ofrece la probabilidad de que la hipótesis nula sea falsa, sólo facilita la probabilidad que tiene un resultado de ocurrir bajo la hipótesis nula pero no indica si dicho resultado ocurre o no ocurre bajo dicha hipótesis. El valor de *p* no tiene relación directa con la presencia o con la magnitud de efectos del tratamiento (Bracey, 1991; Cohen, 1994; Rosenthal, 1993). Además, la prueba estadística es función del tamaño del efecto y del tamaño de la muestra.

3. El valor *p* no informa de la significación sustantiva.

Las pruebas de significación estadística tampoco facilitan información sobre la magnitud o la importancia de un resultado (Grant, 1962; Rosenthal, 1993; Shaver, 1993), aspecto conocido como *significación práctica* (Kirk, 1996) o lo que Levy (1967) denominó *significación sustantiva* y que dentro del área de la psicología clínica se conoce como *significación clínica* (Jacobson, Follette, & Revenstorf, 1984) ¿Cuántos tipos de significación existen? Para realizar interpretaciones de efecto importante, práctico o clínico hay que recurrir al juicio del investigador sobre costos y beneficios tal y como se supone plantea en su hipótesis de investigación. Un juicio orientativo de significación práctica o cambio importante lo encontramos en los valores de Cohen (1990) sobre tamaño del efecto *pequeño* (d=0.2), *mediano* (d=0.5) y *grande* (d=0.8).

A un grupo de 132 alumnos de Psicología formados en las técnicas de significación estadística se les planteaó un diseño experimental con dos condiciones experimentales de 150 sujetos cada una. Se les preguntó "si el valor de *p* obtenido hubiese sido de 0.04 ó quizás de 0.06 ó fuese de 0.0005 podríamos concluir respectivamente..." Como se observa en la gráfica, después de recibir formación de la técnica estadística durante dos años en diferentes asignaturas de análisis de datos, psicometria y diseños de investigación casi el 50% de los alumnos no interpretan correctamente un valor de probabilidad respecto al mágico 0.05 que por consenso establece el límite de la decisión dicotómica estadísticamente significativo/estadísticamente no significativo.



1. Cercano a la significación estadística (p=0.04), no del todo estadísticamente significativo (p=0.06) y muy significativo estadísticamente (p=0.0005).
2. Estadísticamente significativo los valores de p=0.04 y p=0.0005 y estadísticamente no significativo el valor de p=0.06.
3. Marginalmente estadísticamente significativo los valores de p=0.04 y p=0.06 y estadísticamente significativo el valor de p=0.0005.
4. No se puede responder a la pregunta.

Análisis del efecto de tratamiento

Diferentes métodos cuantitativos del cambio terapéutico sustantivo o de análisis de la eficacia del tratamiento permiten analizar si el valor obtenido cumple con el criterio estipulado por la hipótesis de investigación o si el nivel de funcionamiento del sujeto después de la terapia se encuentra dentro de los límites 'normales'. Dentro de la primera orientación destacamos la técnica del meta-análisis donde la estimación del tamaño del efecto medio proporciona el índice de referencia (Glass, 1976; Hedges y Olkin, 1985; Hunter, Schmidt y Jackson, 1982; Rosenthal y Rubin, 1978) y el test de efectos mínimos (Murphy y Myors, 1999). Como técnicas representativas de puntuaciones dentro de la normalidad nos encontramos con el método de comparaciones normativas (Kendall y Grove, 1988) y el Índice del Cambio Fiable (Reliable Change Index, RCI, Jacobson, Follette & Revenstorf, 1984; Jacobson y Truax, 1991; Jacobson, Roberts, Berns y McGlinchey, 1999; Ogles, Lambert y Masters, 1996). En este trabajo nos centramos en las dos primeras técnicas.

Tamaño del efecto y estudios de meta-análisis

El cálculo del tamaño del efecto es una alternativa de análisis cuya información complementa los resultados del procedimiento de significación estadística (American Psychological Association, 1994, 2001; Abelson, 1995; Carver, 1978; 1993; Friedman, 1968; Loftus, 1993; Schafer, 1993; Thompson, 1994, 1996). Detallar el tamaño del efecto en los estudios individuales permite el desarrollo válido de trabajos de meta-análisis y además favorece que el investigador formule hipótesis cada vez más precisas. Sin embargo, conviene recordar que un tamaño del efecto grande no garantiza la importancia teórica o práctica del hallazgo que sólo puede estar respaldada por las hipótesis de investigación y la valoración o juicio del investigador. El problema es decidir en la fase de planificación del diseño qué valor del tamaño del efecto es el apropiado o qué valor se considera mínimo según el área concreta de estudio. Los estudios de meta-análisis ayudan al proceso de planificación proporcionando un tamaño del efecto medio. Dentro del campo clínico el problema es conocer qué valores determinan la normalidad de un constructo psicológico.

El uso del cálculo del tamaño del efecto va aumentando poco a poco especialmente provocado por la política editorial de las revistas que siguen las indicaciones de la A.P.A. pero aún no ha llegado a ser integrado en la interpretación sustantiva de los resultados de investigación, prevaleciendo los datos de significación estadística (Kirk, 1996; Thompson & Snyder, 1997, 1998; Vacha-Haase, Nilsson, Reetz, Lance & Thompson, 2000). Por ello en la quinta edición del manual del A.P.A. (2001) se señala que será *necesario* que siempre se incluya un índice de tamaño del efecto o de magnitud de la relación en la sección de *Resultados*.

"For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section. You can estimate the magnitude of effect or the strength of the relationship with a number of common effect size estimates... The general principle to be followed... is to provide the reader not only with information about statistical significance but also with enough information to assess the magnitude of the observed effect or relationship". (pp. 25-26)

Robinson y Levin (1997) recomiendan utilizar el siguiente criterio. Primero se determina con NHST si el efecto es estadísticamente significativo. Si lo es, se informa del tamaño del efecto. Por su parte Fowler (1985) señala que se puede proceder a la inversa. Primero se comprueba si el tamaño del efecto tiene el valor mínimo que se plantea

en la hipótesis y si es así se pasa a contrastar su significación estadística.

Comprobación estadística de efectos mínimos

Se basa en determinar primeramente el tamaño del efecto mínimo que se considera sustantivo y comprobar posteriormente con NHST la hipótesis de que el efecto especificado es (o no es) estadísticamente significativo, eliminando la interpretación de efectos que no cumplen con el criterio mínimo sustantivo especificado en el plantemiento teórico del estudio dado que son considerados como efectos triviales o sin importancia práctica. Desde este planteamiento se ofrece una respuesta analítica a las críticas relacionadas con la especificación de tamaño de efecto cero, diferencia de medias igual a cero o correlación entre variables igual a cero que el modelo tradicional de contraste de hipótesis plantea con sus *nil hipótesis*. El problema con el que se encuentra el investigador es decidir a priori qué entiende por efecto mínimo cuantificando su valor y así poder determinar el tamaño del efecto que se considerará trivial en el análisis estadístico. Una posible solución puede ser utilizar los valores estandars de Cohen o mejor aún, utilizar los resultados de los trabajos de meta-análisis que ofrecen el tamaño del efecto medio dentro del área concreta objeto de estudio.

El planteamiento que orienta la perspectiva del test de efectos mínimos parte de la comprobación estadística de que la intervención tiene un efecto igual a o menor que un valor mínimo fijado a priori. Si, por ejemplo, el investigador decide que la psicoterapia que explica el 1% de la varianza tiene un efecto excesivamente pequeño como para que sea útil o practico su uso, entonces su tarea consistirá en desarrollar un test estadístico que determine cuán grande ha de ser un tamaño del efecto en una muestra particular para que explique menos del 1% de la varianza. Ese test será una razón *F*, determinando el valor necesario para rechazar la hipótesis que el efecto en la población es igual o menor que el fijado.

El test de efectos mínimos se basa en la distribución *F* no centrada y para desarrollar las tablas con las que testar la hipótesis nula de efectos mínimos, el investigador deberá a) desarrollar la definición operacional de tamaño mínimo; b) calcular el tamaño del parámetro de no centralidad para tal efecto mínimo y c) tabular la correspondiente distribución de *F* no centrada. En este planteamiento se nos presentan dos problemas. Primero *cómo* definir de manera sensata el "efecto mínimo". Segundo, *cómo* generar las tablas de *F* no centrada.

La definición de *qué es el efecto mínimo* requiere un juicio de valor que escapa al análisis estadístico y es previo. Es un problema dependiente de la teoría, de la praxis o del consenso. Otras veces la determinación del tamaño de efecto mínimo puede que exija un análisis de utilidad o de costos y de relevancia social o clínica o cualquier otro criterio que se estime oportuno. Siempre y en cualquier caso, habrá que justificarlo y definirlo y no dar por supuesto como un "prejuicio" dogmáticamente establecido que ese valor ha de ser siempre el valor cero. Por lo tanto, dependerá del área a investigar y de otras concomitancias y también de criterios estadísticos porque conforme se defina un tamaño del efecto mínimo mayor entonces decrecerá la potencia del modelo general lineal para detectar tal efecto, es decir aumenta por tanto la probabilidad de cometer error Tipo II (no detectar un efecto que realmente sí existe). Por lo tanto, antes de llevar a cabo la investigación habrá que delimitar cuál es la relación de riesgo que se acepta respecto de los errores estadísticos de Tipo I (detectar un efecto que realmente no existe) y de Tipo II (no detectar un efecto que realmente sí existe). La relación correcta se puede suponer que es de uno a cuatro como afirma Cohen (1988) o se puede suponer mayor o menor según las circunstancias concretas como por ejemplo el área de investigación.

La generación de las tablas de *F* no centrada es cada día más fácil gracias a las estimaciones que se pueden realizar con las funciones que aparecen en programas como el SPSS o el Excel. También se pueden utilizar programas específicos que computan las probabilidades (Narula & Weistroffer, 1968). Una aproximación basada en la distribución central de la prueba *F* (Horton, 1978; Murphy & Myors, 1999; Patnaik, 1949; Tiku & Yip, 1978) se puede realizar multiplicando el valor crítico de la *F* central, con grados de libertad *g* y *v2*, por el valor *k*. De este modo,

$$\boldsymbol{Pr(F_{v1,v2,\lambda} > c \cdot k)} \text{ es aproximadamente igual a } \boldsymbol{Pr(F_{g,v2} > c)}$$

siendo *Pr* la probabilidad, $F_{v1,v2,\lambda}$ es un valor de la distribución *F* no central, $F_{g,v2}$ es un valor de la distribución *F* central y *c* es una constante. El cómputo de *g* y *k* es el siguiente:

$$g = \frac{(v1+\lambda)^2}{(v1+2\lambda)} \qquad k = \frac{(v1+\lambda)}{v1}$$

El valor de lambda puede ser estimado por medio del porcentaje de varianza explicada, PV, y los grados de libertad del error, v2:

$$\lambda = \frac{(v2 \cdot PV)}{1 - PV}$$

Por ejemplo en las tablas 1 y 2 se detallan los valores de la *F nocentral* obtenidos con el paquetes estadístico SPSS y el Excel. Las tablas proporcionan los valores críticos necesarios para rechazar una hipótesis nula de efectos mínimos de varianza explicada del 1% o menos que en términos del estadístico *d* de Cohen (1988) estamos hablando de tamaños del efecto *pequeño*. Del mismo modo podemos calcular los valores para efectos mínimos de varianza explicada del 5% vinculándose con valores de *d* de efectos entre *pequeños* y *medianos*. Y así para cualquier valor que especifique el criterio de la hipótesis de investigación.

Conclusiones

Un análisis de las aportaciones de los investigadores dedicados al tema de la significación práctica o sustantiva conduce a una conclusión aceptada unánimamente: *se ha abusado de las pruebas de significación estadística aplicándose a todas las situaciones*. Muchas creencias estadísticas deben ser modificadas. El debate acerca de NHST está sin resolver, pero de lo que no cabe duda es que el juicio del investigador es indispensable en la interpretación de los resultados impidiendo un uso mecánico de NHST.

Mientras la polémica sigue su curso durante más de 70 años, el procedimiento de significación de la hipótesis nula sigue siendo la técnica más utilizada por los investigadores y como ya apuntaba Carver (1978), todas las críticas hacia las pruebas de significación estadística han tenido poco efecto. Como consecuencia, es necesario realizar una reflexión sobre su uso adecuado y sobre la formación académica que el futuro profesional recibe en metodología.

Actualmente la cuantificación del impacto del tratamiento, entendido como tamaño del efecto que provoca cambio práctico o formulado como cambio clínico, ha cobrado un papel destacado entre las nuevas formulaciones del análisis estadístico. La técnica del análisis de datos debe adaptarse a las formulaciones sustantivas de cambio práctico o clínico del diseño de investigación. La estimación del tamaño del efecto junto con los datos de significación estadística permiten una comprensión sustantiva de los resultados obtenidos, evitando interpretaciones difíciles de replicar debido a la falta de consistencia del impacto del tratamiento.

La búsqueda del análisis válido del cambio requiere computar y analizar el cambio sustantivo, práctico o clínico. Poder contrastar hipótesis nulas con efecto distinto a cero (hipótesis "non-nil nulls") enriquecera nuestras teorías psicológicas, avanzado el conocimiento y eliminando ciertas polémicas sobre la trivialidad de testar hipótesis con efecto cero, evitando al mismo tiempo la interpretación de resultados estadísticamente significativos sin importancia práctica

Si la objetividad de su planteamiento y su facilidad de cálculo son algunas de las razones del por qué los investigadores siguen adheridos al procedimiento de significación de la hipótesis nula entonces el planteamiento del *análisis de los efectos mínimos* cuenta con esas mismas ventajas, además se evitaría que un resultado no

estadísticamente significativo simplemente lo fuera al aumentar el tamaño de la muestra (aspecto duramente criticado con la aplicación de las pruebas tradicionales de contraste de hipótesis de efecto cero). Sin embargo, el esfuerzo teórico es mayor porque no es suficiente con encontrar cualquier diferencia o cualquier relación sino que la búsqueda se centra en una diferencia o una relación concreta que el investigador debe justificar previamente. La estimación del tamaño del efecto y los resultados de los estudios de meta-análisis ayudarán a los planteamientos de tamaños de efectos.

Cuando se trata de resolver cuestiones prácticas de estimación de tamaños de efectos del tratamiento la técnica del meta-análisis proporciona la información necesaria para el planteamiento teórico del diseño de investigación y el procedimiento estadístico de efectos mínimos permite resolver el análisis del tamaño del efecto especificado.

Tabla 1. Valores críticos de F para comprobar hipótesis de efectos que explican el 1% o menos de varianza obtenidos con alpha= 0.05, v1 como grados de libertad entre y v2 como grados de libertad del error, estimados con el programa Excel.

	v1															
v2	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	60	120
1	163.045	200.487	216.419	225.139	230.616	234.374	237.102	239.180	240.808	242.121	244.106	246.112	248.138	250.179	251.206	252.237	253.275
2	18.879	19.188	19.291	19.342	19.373	19.393	19.408	19.419	19.427	19.434	19.445	19.455	19.465	19.475	19.480	19.486	19.491
3	10.429	9.694	9.368	9.185	9.067	8.985	8.924	8.878	8.841	8.812	8.766	8.720	8.673	8.625	8.601	8.576	8.551
4	8.014	7.082	6.678	6.451	6.306	6.204	6.129	6.071	6.025	5.988	5.931	5.873	5.814	5.753	5.723	5.692	5.660
5	6.935	5.929	5.499	5.256	5.100	4.991	4.910	4.848	4.799	4.758	4.697	4.634	4.569	4.503	4.469	4.435	4.400
6	6.343	5.296	4.851	4.601	4.439	4.326	4.242	4.178	4.126	4.084	4.020	3.954	3.886	3.816	3.780	3.744	3.707
7	5.979	4.902	4.447	4.192	4.027	3.911	3.825	3.758	3.705	3.662	3.595	3.527	3.456	3.384	3.346	3.308	3.269
8	5.739	4.636	4.174	3.914	3.746	3.628	3.540	3.472	3.418	3.374	3.306	3.235	3.163	3.088	3.049	3.009	2.969
9	5.573	4.446	3.977	3.714	3.544	3.424	3.335	3.266	3.210	3.165	3.096	3.024	2.950	2.872	2.832	2.791	2.750
10	5.456	4.306	3.831	3.564	3.392	3.270	3.180	3.110	3.054	3.008	2.937	2.864	2.788	2.708	2.667	2.625	2.582
11	5.372	4.199	3.718	3.448	3.274	3.151	3.059	2.988	2.931	2.885	2.813	2.738	2.661	2.580	2.538	2.495	2.450
12	5.311	4.116	3.629	3.356	3.180	3.055	2.963	2.891	2.833	2.786	2.713	2.638	2.559	2.476	2.433	2.389	2.343
13	5.268	4.050	3.557	3.281	3.103	2.978	2.884	2.811	2.753	2.705	2.632	2.555	2.475	2.391	2.347	2.302	2.255
14	5.238	3.998	3.498	3.220	3.040	2.914	2.819	2.745	2.687	2.638	2.564	2.486	2.404	2.319	2.274	2.228	2.180
15	5.218	3.956	3.450	3.169	2.987	2.860	2.764	2.690	2.630	2.581	2.506	2.427	2.345	2.258	2.212	2.165	2.117
16	5.206	3.922	3.410	3.126	2.943	2.814	2.717	2.642	2.582	2.533	2.457	2.377	2.294	2.205	2.159	2.111	2.062
17	5.200	3.894	3.376	3.089	2.905	2.774	2.677	2.602	2.541	2.491	2.414	2.334	2.249	2.160	2.113	2.064	2.013
18	5.200	3.871	3.348	3.058	2.872	2.740	2.642	2.566	2.505	2.455	2.377	2.296	2.210	2.120	2.072	2.023	1.971
19	5.205	3.853	3.323	3.031	2.843	2.711	2.612	2.535	2.473	2.423	2.344	2.262	2.176	2.084	2.036	1.986	1.933
20	5.213	3.839	3.303	3.008	2.818	2.685	2.585	2.508	2.445	2.394	2.315	2.232	2.145	2.053	2.004	1.953	1.899
21	5.224	3.827	3.285	2.988	2.796	2.662	2.561	2.483	2.421	2.369	2.289	2.206	2.118	2.024	1.975	1.923	1.869
22	5.238	3.818	3.270	2.970	2.777	2.642	2.540	2.462	2.399	2.347	2.266	2.182	2.093	1.999	1.949	1.896	1.841
23	5.254	3.812	3.258	2.955	2.760	2.624	2.522	2.442	2.379	2.327	2.245	2.161	2.071	1.975	1.925	1.872	1.816
24	5.272	3.807	3.247	2.941	2.745	2.608	2.505	2.425	2.361	2.308	2.227	2.141	2.051	1.954	1.903	1.850	1.793
25	5.292	3.804	3.238	2.929	2.732	2.593	2.490	2.409	2.345	2.292	2.210	2.123	2.032	1.935	1.883	1.829	1.772
26	5.313	3.803	3.230	2.919	2.720	2.580	2.476	2.395	2.330	2.277	2.194	2.107	2.015	1.917	1.865	1.810	1.753
27	5.335	3.802	3.224	2.910	2.709	2.569	2.464	2.382	2.317	2.263	2.180	2.092	2.000	1.901	1.848	1.793	1.735
28	5.359	3.803	3.219	2.902	2.700	2.558	2.453	2.371	2.305	2.251	2.167	2.079	1.986	1.886	1.833	1.777	1.718
29	5.384	3.805	3.215	2.895	2.692	2.549	2.443	2.360	2.294	2.239	2.155	2.066	1.973	1.872	1.818	1.762	1.702
30	5.410	3.808	3.212	2.889	2.684	2.540	2.433	2.350	2.284	2.229	2.144	2.055	1.960	1.859	1.805	1.748	1.688
40	5.702	3.872	3.213	2.864	2.643	2.490	2.376	2.288	2.217	2.160	2.070	1.975	1.875	1.767	1.710	1.648	1.582
50	6.031	3.970	3.250	2.874	2.638	2.475	2.355	2.262	2.187	2.126	2.032	1.933	1.828	1.715	1.654	1.589	1.518
60	6.378	4.086	3.304	2.900	2.650	2.477	2.350	2.253	2.175	2.111	2.012	1.909	1.800	1.682	1.618	1.550	1.475
70	6.734	4.211	3.367	2.936	2.671	2.489	2.356	2.253	2.172	2.105	2.003	1.895	1.782	1.660	1.593	1.522	1.443
80	7.097	4.343	3.437	2.978	2.698	2.506	2.367	2.260	2.175	2.106	1.999	1.888	1.771	1.644	1.575	1.502	1.420
90	7.464	4.478	3.509	3.024	2.728	2.528	2.382	2.271	2.182	2.110	2.000	1.885	1.764	1.633	1.562	1.486	1.402
100	7.833	4.616	3.585	3.072	2.762	2.552	2.400	2.284	2.192	2.117	2.003	1.884	1.759	1.625	1.553	1.474	1.387
120	8.577	4.896	3.742	3.174	2.834	2.606	2.441	2.316	2.217	2.137	2.015	1.889	1.757	1.616	1.540	1.457	1.365
150	9.702	5.326	3.984	3.335	2.950	2.694	2.511	2.372	2.263	2.176	2.042	1.905	1.763	1.611	1.530	1.442	1.344
200	11.587	6.052	4.399	3.613	3.154	2.852	2.637	2.476	2.351	2.250	2.098	1.943	1.784	1.616	1.527	1.431	1.324
300	12.013	6.547	5.243	4.185	3.577	3.183	2.906	2.700	2.542	2.415	2.226	2.036	1.844	1.646	1.541	1.431	1.308
400	14.971	7.829	5.555	4.451	3.801	3.374	3.071	2.933	2.741	2.589	2.363	2.138	1.913	1.684	1.566	1.441	1.305
500	15.605	8.304	6.333	4.994	4.212	3.701	3.341	3.073	2.867	2.703	2.457	2.243	1.986	1.727	1.594	1.456	1.307
600	18.181	9.476	6.643	5.252	4.430	4.029	3.612	3.303	3.065	2.877	2.596	2.318	2.041	1.771	1.625	1.474	1.312
1000	24.231	12.540	8.680	6.768	5.804	5.006	4.443	4.023	3.700	3.443	3.059	2.711	2.321	1.945	1.748	1.548	1.341
5000	76.169	38.569	26.039	19.901	16.117	13.595	11.796	10.504	9.447	8.603	7.338	6.102	4.853	3.605	2.975	2.347	1.716
10000	135.141	68.063	45.825	34.615	27.889	23.407	20.254	17.846	15.974	14.476	12.257	10.007	7.788	5.549	4.425	3.310	2.188

Tabla 2. Valores críticos de F para comprobar hipótesis de efectos que explican el 1% o menos de varianza obtenidos con lambd=0.05, v1 como grados de libertad entre y v2 como grados de libertad del error, estimados con el paquete estadístico SPSS.

	v1															
v2	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	60	120
1	163.085	200.509	216.434	225.151	230.627	234.380	237.110	239.184	240.813	242.126	244.111	246.116	248.138	250.179	251.207	252.238	253.274
2	18.887	19.192	19.293	19.344	19.374	19.395	19.409	19.420	19.428	19.435	19.445	19.455	19.465	19.476	19.481	19.486	19.491
3	10.434	9.697	9.370	9.186	9.068	8.986	8.925	8.879	8.842	8.812	8.767	8.720	8.673	8.625	8.601	8.576	8.552
4	8.018	7.084	6.680	6.453	6.307	6.205	6.129	6.072	6.026	5.988	5.932	5.874	5.814	5.754	5.723	5.692	5.660
5	6.938	5.932	5.500	5.258	5.101	4.992	4.911	4.849	4.799	4.75v9	4.697	4.634	4.570	4.503	4.469	4.435	4.400
6	6.346	5.298	4.853	4.602	4.440	4.327	4.243	4.178	4.127	4.085	4.020	3.954	3.886	3.816	3.780	3.744	3.707
7	5.980	4.904	4.449	4.193	4.028	3.911	3.825	3.759	3.706	3.662	3.596	3.527	3.457	3.384	3.346	3.308	3.269
8	5.739	4.637	4.175	3.915	3.747	3.629	3.541	3.473	3.419	3.374	3.306	3.236	3.163	3.088	3.049	3.009	2.969
9	5.572	4.448	3.979	3.715	3.545	3.425	3.335	3.266	3.211	3.166	3.096	3.024	2.950	2.872	2.832	2.791	2.750
10	5.453	4.307	3.832	3.565	3.393	3.271	3.181	3.110	3.054	3.008	2.937	2.864	2.788	2.709	2.668	2.625	2.582
11	5.367	4.200	3.719	3.449	3.275	3.152	3.060	2.989	2.932	2.885	2.813	2.739	2.661	2.580	2.538	2.495	2.450
12	5.304	4.117	3.630	3.357	3.181	3.056	2.964	2.892	2.834	2.787	2.714	2.638	2.559	2.476	2.433	2.389	2.343
13	5.259	4.051	3.558	3.283	3.104	2.979	2.885	2.812	2.754	2.706	2.632	2.555	2.475	2.391	2.347	2.302	2.255
14	5.226	3.998	3.500	3.221	3.041	2.915	2.820	2.746	2.687	2.639	2.564	2.486	2.405	2.319	2.274	2.228	2.180
15	5.204	3.956	3.451	3.170	2.989	2.861	2.765	2.691	2.631	2.582	2.507	2.428	2.345	2.258	2.213	2.166	2.117
16	5.189	3.921	3.411	3.127	2.944	2.815	2.718	2.643	2.583	2.534	2.457	2.378	2.294	2.206	2.159	2.111	2.062
17	5.180	3.893	3.377	3.091	2.906	2.775	2.678	2.602	2.542	2.492	2.415	2.334	2.249	2.160	2.113	2.064	2.014
18	5.177	3.870	3.348	3.059	2.873	2.741	2.643	2.567	2.506	2.455	2.377	2.296	2.211	2.120	2.072	2.023	1.971
19	5.178	3.851	3.324	3.032	2.844	2.712	2.613	2.536	2.474	2.423	2.345	2.263	2.176	2.084	2.036	1.986	1.933
20	5.183	3.836	3.303	3.009	2.819	2.686	2.586	2.509	2.446	2.395	2.316	2.233	2.146	2.053	2.004	1.953	1.900
21	5.190	3.824	3.286	2.989	2.798	2.663	2.562	2.484	2.422	2.370	2.290	2.206	2.118	2.024	1.975	1.923	1.869
22	5.200	3.814	3.271	2.971	2.778	2.643	2.541	2.463	2.399	2.348	2.267	2.183	2.094	1.999	1.949	1.896	1.841
23	5.212	3.807	3.258	2.956	2.761	2.625	2.523	2.443	2.380	2.327	2.246	2.161	2.071	1.976	1.925	1.872	1.816
24	5.226	3.801	3.247	2.942	2.746	2.609	2.506	2.426	2.362	2.309	2.227	2.142	2.051	1.955	1.903	1.850	1.793
25	5.241	3.797	3.237	2.930	2.733	2.594	2.491	2.410	2.346	2.293	2.210	2.124	2.033	1.935	1.884	1.829	1.772
26	5.258	3.795	3.229	2.920	2.721	2.581	2.477	2.396	2.331	2.278	2.195	2.108	2.016	1.918	1.865	1.811	1.753
27	5.276	3.794	3.223	2.911	2.710	2.570	2.465	2.383	2.318	2.264	2.181	2.093	2.000	1.901	1.849	1.793	1.735
28	5.295	3.794	3.217	2.902	2.701	2.559	2.454	2.372	2.306	2.252	2.168	2.079	1.986	1.886	1.833	1.777	1.718
29	5.315	3.795	3.213	2.896	2.692	2.550	2.444	2.361	2.295	2.240	2.156	2.067	1.973	1.872	1.819	1.762	1.702
30	5.335	3.797	3.209	2.889	2.685	2.541	2.434	2.351	2.285	2.230	2.145	2.055	1.961	1.859	1.805	1.748	1.688
40	5.570	3.848	3.207	2.862	2.643	2.491	2.377	2.289	2.218	2.161	2.070	1.976	1.876	1.768	1.710	1.648	1.582
50	5.828	3.931	3.238	2.869	2.637	2.475	2.355	2.263	2.188	2.127	2.033	1.934	1.829	1.716	1.654	1.589	1.518
60	6.094	4.029	3.285	2.893	2.647	2.476	2.350	2.253	2.175	2.112	2.013	1.910	1.801	1.682	1.618	1.550	1.475
70	6.360	4.133	3.341	2.925	2.666	2.487	2.355	2.253	2.172	2.106	2.004	1.896	1.783	1.660	1.594	1.522	1.444
80	6.624	4.241	3.401	2.963	2.690	2.503	2.365	2.259	2.175	2.106	2.000	1.889	1.772	1.645	1.576	1.502	1.420
90	6.885	4.351	3.463	3.003	2.718	2.523	2.379	2.269	2.182	2.110	2.000	1.886	1.764	1.634	1.563	1.487	1.402
100	7.142	4.461	3.528	3.046	2.748	2.545	2.396	2.282	2.191	2.117	2.003	1.885	1.760	1.626	1.553	1.475	1.387
120	7.645	4.681	3.660	3.135	2.813	2.594	2.434	2.312	2.215	2.136	2.015	1.890	1.758	1.617	1.540	1.458	1.366
150	8.370	5.006	3.859	3.274	2.916	2.674	2.498	2.364	2.258	2.172	2.041	1.905	1.764	1.612	1.531	1.443	1.344
200	9.512	5.533	4.188	3.507	3.093	2.814	2.613	2.460	2.340	2.242	2.094	1.942	1.785	1.617	1.528	1.432	1.324
300	11.616	6.528	4.823	3.965	3.447	3.100	2.850	2.661	2.514	2.395	2.214	2.031	1.843	1.646	1.543	1.432	1.309
400	13.554	7.464	5.428	4.407	3.792	3.381	3.086	2.863	2.690	2.550	2.340	2.126	1.909	1.684	1.567	1.443	1.306
500	15.382	8.355	6.009	4.834	4.127	3.655	3.317	3.063	2.865	2.705	2.465	2.223	1.978	1.726	1.595	1.458	1.308
600	17.128	9.212	6.570	5.248	4.454	3.924	3.544	3.259	3.037	2.859	2.591	2.321	2.048	1.769	1.625	1.475	1.313
1000	23.611	12.417	8.685	6.818	5.698	4.951	4.417	4.017	3.705	3.455	3.080	2.704	2.327	1.945	1.751	1.552	1.342
5000	77.142	39.122	26.449	20.112	16.310	13.775	11.965	10.607	9.551	8.706	7.438	6.171	4.903	3.635	3.001	2.365	1.726
10000	137.394	69.234	46.514	35.154	28.338	23.794	20.549	18.114	16.221	14.706	12.434	10.162	7.890	5.618	4.481	3.345	2.206

Referencias

Abelson, R.P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Erlbaum

American Psychological Association (1994) Publications manual of the *American Psychological Association (4th ed.).* Washington, DC: Author.

American Psychological Association (2001) Publications manual of the *American Psychological Association (5th ed.).* Washington, DC: Author.

Bracey, G.W. (1991). Sense, non-sense, and statistics. Phi Delta *Kappan, 73,* 335.

Carver, R.P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48,* 378-399.

Carver, R.P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education, 61,* 287-292.

Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304-1312.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49,* 997-1003.

Fowler, R. L. (1985). Testing for substantive significance in applied research by specifying nonzero null hypotheses. *Journal of Applied Psychology, 70,* 215-218.

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin, 70,* 245-251.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 5,*3-8.

Grant, D.A. (1962). Testing the null hypothesis and the strategy and tactics on investigating theoretical models. *Psychological Review, 69,* 54-61.

Hedges, L. V. y Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Horton, R. L. (1978). *The general linear model: Data analysis in the social and behavioral sciences.* New York, NY: McGraw-Hill.

Hunter, J. E., Schmidt, F. L. y Jackson, G. B. (1982). *Meta-analysis: cumulating findings across research.* Beverly Hills, CA: Sage.

Jacobson, N.S. y Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19.

Jacobson, N.S., Follette, W.C. y Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Journal of Consulting and Clinical Psychology, 15,* 336-352.

Jacobson, N.S., Roberts, L.J., Berns, S.B. y McGlinchey, J.B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application and alternatives. *Journal of Consulting and Clinical Psychology, 67*(3), 300-307.

Kendall, P.C. y Grove, W.M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment, 10,* 147-158.

Kendall, P.C., Flannery-Schroeder , E.C. Ford, J.D. (1999).Therapy outcome research methods. En P.C. Kendall, J.N. Butcher y G.N. Holmbeck (eds.). *Handbook of research methods in clinical psychology.* New York: Wiley and Sons.

Kirk, R. E. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement, 56,*746-759.

Levy, P. (1967). Substantive significance of significant differences between two groups. *Psychological Bulletin, 67,* 37-40.

Loftus, G.R. (1993). Editorial comment. *Memory & Cognition, 21,* 1-3.

Murphy, K. R. & Myors, B. (1999). Testing the hypothesis that treatments have negligible effects: Minimun-effect tests in the general linear model. *Journal of Applied Psychology, 84,* 234-2484.

Narula, S., & Weistroffer, H. (1968). Computation of probability and noncentrality parameter of a noncentral F distribution. *Communications in Statistics B, 15,* 871-878.

Ogles, B.M., Lambert, M.J. y Masters, K.S. (1996). *Assessing outcome in clinical practice.* Boston: Allyn and Bacon.

Patnaik, P.B. (1949). The non-central c??and F-distributions and their applications. *Biometrika, 36,* 202-232.

Robinson, D., y Levin, J. (1997). Reflections on statistical and substantive significance with a slice of replication. *Educational Researcher, 26,* 21-26.

Rosenthal, R y Rubin, D. (1978). Interpersonal expectancy effects: the first 345 studies. *Statistical Science, 3,* 120-125.

Rosenthal, R. (1993). Cumulative evidence. En G. Keren & C. Lewis (Eds.). *A handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale, NJ: Erlbaum.

Shaver, W.D. (1993). Interpreting statistical significance and nonsignificance. *Journal of Experimental Education, 61,* 383-387.

Thompson, B. & Snyder, P.A. (1997). Statistical significance testing practices in the Journal of Experimental Education. *Journal of Experimental Education, 66,* 75-83.

Thompson, B. & Snyder, P.A. (1998). Statistical significance and reliability analyses in recent JCD research articles. *Journal of Counseling and Development, 76,* 436-441.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 26,* 29-32.

Thompson, B. (1996). Guidelines for authors. *Educational and Psychological Measurement, 54,* 837-847.

Tiku, M.L. & Yip, D.Y.N. (1978). A four-moment approximation based on the F distribution. *Australian Journal of Statistics, 20,* 257-261.

Vacha-Haase, T., Nilsson, J.E., Reetz, D.R., Lance, T.S. & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology, 10,* 413-425.