

## Un estudio de la potencia estadística en Anales de Psicología (1984-1991)

Julio Sánchez Meca <sup>1</sup>

Antonio Valera Espín

Antonio P. Velandrino Nicolás

Fulgencio Marín Martínez

Universidad de Murcia

**Resumen.** Con objeto de conocer la potencia estadística de las investigaciones publicadas en *Anales de Psicología* a lo largo de su existencia (1984-1991), se ha analizado la potencia de las 16 investigaciones publicadas en la revista susceptibles de este tipo de cálculo. Los resultados no parecen diferir de los originalmente obtenidos por Cohen (1962), mostrando potencias medias de .13, .47 y .76 para tamaños del efecto bajos, medios y altos, respectivamente. Además, la potencia media para los TEs estimados desde los propios estudios alcanzó un valor de .71, próximo al mínimo de .80 recomendado por Cohen (1988). Por último, estos resultados son comparados con otros estudios de potencia recientes.  
**Palabras clave:** Potencia estadística; tamaño del efecto; contraste de hipótesis; nivel de significación.

**Abstract.** The aim of this paper was to determine the statistical power of the research published in the journal *Anales de Psicología* across its life (1984-1991). The sixteen studies available for this calculation were used for analyzing their statistical power. The results do not seem to differ from those originally obtained by Cohen (1962), showing average powers of .13, .47, and .76 for small, medium, and large effect sizes, respectively. Also, the mean power for the estimated effect sizes from the proper studies increased to .71, very close to the minimum of .80 recommended by Cohen (1988). Finally, the results are compared to other recent power studies.

**Key words:** Statistical power; effect size; hypothesis testing; significance level.

### Introducción

Uno de los objetivos prioritarios –si no el que más– en Ciencias Sociales y del Comportamiento consiste en descubrir relaciones entre variables. Esta circunstancia plantea la necesidad de que el investigador dedique especial atención a la posibilidad de detectar la presencia de tales relaciones. O, dicho de otra forma, resulta preciso que el investigador se asegure de que maneja una alta probabilidad de rechazar adecuadamente la hipótesis nula. Esta probabilidad es conocida con el nombre de *potencia* de la prueba estadística (Hays, 1988). Lamentablemente, es un hecho bien conocido (Kazdin y Bass, 1989; Rossi, 1983 abril, 1990; Sedlmeier y Gigerenzer, 1989) la escasa atención que los investigadores sociales y del comportamiento dedican a la evaluación del nivel de potencia de sus estudios. Un reducido grupo de metodólogos sociales, a cuya cabeza podemos situar a Jacob Cohen, vienen llamando la atención desde hace tres décadas sobre la imperiosa necesidad de controlar la potencia de las investigaciones para que las conclusiones obtenidas presenten una alta tasa de fiabilidad.

Una de las razones que evidencia el importante papel que desempeña la potencia estadística en una investigación es el de indicar la viabilidad de realizar la propia investigación. En efecto, de acuerdo con Rossi (1990), "... el conocimiento de la potencia de una prueba estadística indica la posibilidad de obtener un resultado estadísticamente significativo" (p. 646). Así, no sería deseable llevar a cabo una investigación con una baja potencia (es decir, con una baja probabilidad de rechazar correctamente la hipótesis nula); seguramente, resultarían inútiles todos los recursos y esfuerzos movilizados en la ejecución de una investigación en la que hubiese escasa posibilidad de detectar la presencia del efecto estudiado.

En esta línea, con el presente trabajo nos proponemos alentar a los investigadores a que dediquen una mayor atención a la potencia estadística en el convencimiento de que ello redundará en beneficio de la seguridad de sus resultados estadísticos.

<sup>1</sup>*Dirección:* Dr. Julio Sánchez Meca. Dpto. Metodología y Análisis del Comportamiento. Facultad de Psicología. Apdo. 4021, 30080 Murcia (Spain).

©Copyright 1992. Secretariado de Publicaciones e Intercambio Científico. Universidad de Murcia. Murcia (Spain). ISSN: 0212-09728.

En lo que sigue se expone, en primer lugar, una breve presentación del concepto de potencia estadística y su relación con los tipos de error y el tamaño del efecto; a continuación, se revisan los estudios de potencia efectuados en el campo de las Ciencias Sociales y del Comportamiento, y se finaliza con un estudio de la potencia estadística de la revista *Anales de Psicología* a lo largo de su historia.

### Los parámetros del contraste de hipótesis

El objetivo de todo contraste de hipótesis puede resumirse, de una forma un tanto genérica, en probar la significación estadística de la posible relación entre dos o más variables. Para ello se formula una hipótesis nula ( $H_0$ ) en los términos de que no existe relación entre las variables, conjuntamente con una hipótesis alternativa ( $H_1$ ) donde se afirma la veracidad de dicha relación.

La resolución de un contraste de hipótesis culmina, bien con la aceptación, o bien con el rechazo de la  $H_0$ , en base a la probabilidad de los datos muestrales bajo el supuesto de que la  $H_0$  es verdadera (valor de probabilidad  $p$  asociado al estadístico de contraste). Cuando dicho valor  $p$  es inferior a un nivel crítico alfa prefijado por el investigador (generalmente de .05), lo suficientemente pequeño como para desconfiar en la compatibilidad de los datos con la formulación de la  $H_0$ , la decisión que se adopta es la de rechazarla. Se asume con ello el riesgo de cometer un error tipo I, aunque con una probabilidad razonablemente baja, determinada por el parámetro  $\alpha$  (probabilidad de rechazar la  $H_0$  siendo realmente verdadera).

Cuando, por el contrario, la probabilidad  $p$  es superior al nivel  $\alpha$ , decidimos aceptar la  $H_0$ , asumiendo el riesgo de un error tipo II que no se suele valorar, aun a pesar de las consecuencias que como veremos se derivan de esta desatención.

Si la hipótesis alternativa, como de hecho sucede en la investigación empírica, no plantea en términos cuantificables la magnitud de la relación que se pretende probar, resulta imposible controlar a priori la probabilidad  $\beta$  de cometer un error tipo II, de aceptar la  $H_0$  cuando de hecho existe la relación en una magnitud igual o superior a la formulada en la  $H_1$ .

Formulando la  $H_1$  como la existencia de una determinada magnitud o tamaño del efecto (TE) en la relación probada, enriquecemos el planteamiento del contraste de hipótesis desde dos puntos de vista: (a) vamos más allá del reduccionismo que supone la mera decisión dicotómica de afirmar o negar el que las variables estén relacionadas, posibilitando valorar si la relación alcanza la fuerza o relevancia con la que se plantea en la  $H_1$  (rechazamos la  $H_0$ ), o si por el contrario su magnitud no es considerada de interés por ser inferior al TE probado (aceptamos la  $H_0$ ); y (b) por otro lado, somos capaces de conocer el riesgo de cometer un error tipo II, adoptando las medidas oportunas para que dicha probabilidad  $\beta$  no sobrepase ciertos límites razonables que garanticen un mínimo de rigor en nuestro proceder metodológico.

A modo de ejemplo, podríamos estar interesados en estudiar la relación entre las variables "sexo" y "rendimiento en un determinado tipo de tareas", para lo que aplicaríamos una prueba  $t$  de diferencias entre medias. Si en lugar de plantear la  $H_1$  en términos de que ambos sexos difieren significativamente en el rendimiento (sea en favor de los varones o de las mujeres), formulamos la hipótesis de que difieren como mínimo en 0'5 desviaciones típicas, nuestra conclusión tendrá un referente cuantitativo de la medida en que ambas variables se relacionan, además de permitirnos conocer la potencia de la prueba para detectar dicha relación, si realmente existe, en una magnitud igual o superior a 0'5 unidades típicas. Si se acepta la  $H_0$ , sabremos que no existen diferencias, o que si las hay, en todo caso serían inferiores a un TE de 0'5, que conforme a los criterios de Cohen equivaldría a un efecto medio (véase el concepto de TE y su magnitud en Cohen, 1988). Si rechazamos la  $H_0$ , concluiremos que las diferencias son como mínimo de una magnitud media (TE = 0'5) que, según el contexto, pudiera ser considerada relevante.

El tamaño muestral ( $n$ ), el tamaño del efecto (TE), la probabilidad de error tipo I ( $\alpha$ ) y la probabilidad de error tipo II ( $\beta$ ) constituyen los cuatro parámetros a considerar conjuntamente antes de planificar un contraste de hipótesis, a fin de determinar un equilibrio adecuado que permita

minimizar las dos tasas de error. Fijado  $\alpha$  en un nivel del 5%, como de hecho suele plantearse en gran parte de la investigación en Ciencias Sociales, el problema que aquí se suscita es el de garantizar una probabilidad  $\beta$  pequeña, o lo que es lo mismo, una potencia  $(1 - \beta)$  que permita detectar el TE probado, si realmente existe, con una alta probabilidad.

Manteniendo constantes  $n$  y el TE, la relación entre  $\alpha$  y  $\beta$  es inversa, con lo que el aumento de uno de estos valores de probabilidad da lugar a la disminución del otro. Tal y como sugiere Cohen (1990), en determinadas condiciones podríamos optar por incrementar ligeramente la tasa de error tipo I por encima del convencional 5%, reduciendo el tradicionalmente desatendido error tipo II. Sin llegar a tal punto, nosotros proponemos asumir como constante  $\alpha = .05$ ; buscando asimismo que el error tipo II no ascienda por encima del 20% ( $\beta = .20$ ). Con ello garantizamos una potencia mínima del 80%, siguiendo la recomendación, ampliamente aceptada, de Cohen (1988).

Para un determinado TE, que determina la distancia entre el centro de las distribuciones de probabilidad que definen las hipótesis nula y alternativa, existe un  $n$  a partir del cual la potencia alcanza el valor de .80. Tamaños muestrales superiores implicarían un menor solapamiento de ambas distribuciones, y con ello un error tipo II inferior o una mayor potencia.

Fijando un tamaño muestral constante, existirá un TE que podrá ser detectado con una potencia de .80. TEs superiores se asociarán con una potencia mayor, debido de nuevo al menor solapamiento entre las distribuciones.

Por consiguiente, fijado  $\alpha$  en un 5%, planteando el TE que queremos probar y persiguiendo como objetivo el que la potencia ascienda a un 80%, la determinación del tamaño muestral condicionará la adecuación del contraste de hipótesis al rigor metodológico que pretendemos asegurar.

Las consecuencias de no controlar el error tipo II puede llevarnos a encontrar estudios con una potencia para detectar TEs de interés en Ciencias Sociales muy pequeña (inferior al 50%), media (del orden de un 50%) o grande (del orden de un 80%), conforme a los criterios estándar establecidos por autores tales como Cohen (1977) o Rosnow y Rosenthal (1988).

Un estudio que se ha realizado bajo condiciones de baja potencia estadística para detectar un determinado TE relevante, está destinado a errar en sus conclusiones con una alta probabilidad, la cual va a coincidir con la tasa de error tipo II. Será más probable aceptar la hipótesis nula que rechazarla, cuando en realidad existe una relación de interés, que puede quedar almacenada y olvidada en los denominados *file-drawers* o conjunto de investigaciones no publicadas por no alcanzar la significación estadística (Rosenthal, 1979).

Pero aun en el caso de que con una baja potencia hayamos conseguido rechazar la hipótesis nula, resultará difícil replicar este resultado en futuras investigaciones (Rosenthal, 1990). Bajo estas condiciones, se hace muy difícil contar con estudios consistentes en cuanto a la consecución de resultados significativos o no significativos en el estudio de una determinada relación.

Potencias del orden del 50% en relación con un TE que realmente existe y nos interesa detectar como significativo nos conducen a hacer de la inferencia un ejercicio completamente abandonado al azar, donde la probabilidad de aceptar o rechazar la hipótesis nula va a ser similar, cuando la decisión correcta sería el rechazo de ésta. Volverían a darse los mismos problemas mencionados para el caso de que la potencia sea baja: engrosamos los *file-drawers* y de cada dos investigaciones tan sólo en una rechazamos la hipótesis nula.

### Revisión de los estudios de potencia

Desde que en 1962 Cohen dio a conocer su canónico trabajo en el que llevó a cabo un estudio de la potencia estadística de los artículos aparecidos en 1960 en la revista *Journal of Abnormal and Social Psychology*, se han llevado a cabo alrededor de una treintena de investigaciones similares. En la Tabla 1 se presenta un resumen de los resultados obtenidos en los estudios de potencia recopilados en diversas áreas de investigación relacionadas con las ciencias del comportamiento; en estos estudios están implicadas disciplinas tales como la sociología, las ciencias económicas o las ciencias de la

educación.

Tabla I: Resultados obtenidos en los estudios de potencia

Fuente	Revista(s)	Año(s)	Artícs.	Pequeño	Medio	Grande
Cohen (1962)	J. Abn. & Soc. Psych.	1960	70	.18	.48	.83
Brewer (1972)	Amer. Educ. Res. J.	1969-71	47	.14	.58	.78
Jones y Brewer (1972)	Res. Quart.	1969-71	106	.15	.54	.83
Pennick y Brewer (1972)	J. Res. Sci. Teach.	1969-70	66	.22	.71	.87
Brewer y Owen (1973)	J. Educ. Meas.	1969-71	13	.21	.72	.96
Katzer y Sodt (1973)	J. Commun.	1971-72	31	.23	.56	.79
Haase (1974)	Coun. Educ. & Superv.	1968-71	60	.19	.46	.72
Spreitzer y Chase (1974)	Amer. Soc. Rev.	1972-73	34	.55	.84	.94
Chase y Tucker (1975)	Amer. For. Ass. J.	1973	46	.18	.52	.79
	Central States Speech J.	1973				
	J. Commun.	1973				
	Quart. J. Speech	1973				
	Sout. Speech J.	1973				
	Speech Monog.	1973				
	Speech Teacher	1973				
	Today's Speech	1973				
	West. Speech	1973				
Kroll y Chase (1975)	J. Commun. Disab.	1973-74	62	.16	.44	.73
	J. Speech & Hear. Res.	1973-74				
Chase y Baran (1976)	J. Quart.	1974	48	.34	.76	.91
	J. Broad.	1974				
Chase y Chase (1976)	J. Appl. Psychol.	1974	121	.25	.67	.86
Christensen y Christensen (1977)	Res. Quart.	1975	43	.18	.39	.62
Chase et al. (1978)	Amer. J. Phys. Anthr.	1976	25	.38	.62	.81
Levenson (1980)	Gerontologist	1961-77	56	.37	.88	.96
	J. Gerontology	1946-77				
Sawyer y Ball (1981)	J. Market. Res.	1979	23	.41	.89	.98
Ottenbacher (1982)	Amer. J. Occup. Ther.	1980	22	.37	.65	.93
Daly y Hexamer (1983)	Res. in the Teach. Engl.	1978-80	57	.22	.63	.86
Woolley y Dawson (1983)	J. Res. Sci. Teach.	1977-80	192	.23	.63	.85
Woolley (1983)	J. Med. Educ.	1980-82	100	.23	.69	.90
Orme y Combs-Orme (1986)	Soc. Work Res. & Abst.	1977-84	49	.31	.76	.92
Orme y Tolman (1986)	J. Soc. Work Educ.	1976-85	64	.20	.68	.88
Mazen et al. (1987)	Strat. Manag. J.	1982-84	44	.23	.59	.83
	Acad. Manag. J.	1984				
Mazen et al. (1987)	Acad. Manag. J.	1984	84	.31	.77	.91
	J. Manag.	1984				
	Proc. Midw. Div. Ac. Man.	1984				
Sedlmeier y Gigerenzer (1989)	J. Abn. Psychol.	1984	54	.21	.50	.84
Rossi (1990)	J. Abn. Psychol.	1982	221	.17	.57	.83
	J. Cons. & Clin. Psych.	1982				
	J. Pers. & Soc. Psych.	1982				
Total y medias ponderadas			1738	.23	.62	.85

Nota.- A excepción de Cohen (1962) y Sedlmeier y Gigerenzer (1989), los cálculos de potencia que se presentan en la tabla están basados en las definiciones más recientes de TEs pequeño, medio y grande propuestas por Cohen (1977, 1988). Las potencias medias presentadas en la parte inferior se han obtenido ponderando cada estudio por el número de artículos. (Modificado de Rossi, 1990, p. 648.)

Aunque se observa cierta variabilidad, la tendencia general conduce a una interpretación pesimista del estado en que se encuentra la potencia estadística en la investigación social. De hecho, salvo para

TEs grandes, la potencia media de los estudios se sitúa en torno al 50% (para TEs medios) e incluso menos (para TEs pequeños).

Así pues, siguiendo esta línea de trabajo, presentamos a continuación un estudio de la potencia estadística de las investigaciones publicadas en *Anales de Psicología* con objeto de poder compararla con la obtenida en la investigación social, en general, y en la investigación psicológica, en particular.

### Método

Se revisaron todos los artículos publicados en *Anales de Psicología* desde su inicio en 1984 hasta 1991, clasificándolos según las categorías: (a) artículos de revisión o reflexión teórica, es decir, estudios cualitativos o cuantitativos sobre el estado del arte en un tópico determinado y reflexiones teóricas de cualquier índole; (b) artículos que fundamentan su contenido en una investigación empírica; (c) artículos que no se ajustan de una manera clara en ninguna de las dos categorías anteriores (comentarios, actas de reuniones, editoriales, reseñas de libros, etc.). Obviamente, de esta triple taxonomía sólo pudieron incluirse en el estudio de potencia los informes de investigaciones empíricas, por lo que fueron eliminados todos los artículos clasificados en las categorías (a) y (c).

Por otra parte, no todos los estudios empíricos permiten el cálculo de la potencia estadística. Se excluyeron, pues, aquellas investigaciones que no aportaban pruebas estadísticas de contraste (por ejemplo, todas las técnicas de reducción de la dimensionalidad: análisis factorial, escalamiento multidimensional, análisis cluster), o aquellas otras que, aun haciéndolo, no ofrecían los datos suficientes como para realizar los cálculos oportunos; también quedaron excluidos los diseños de caso único.

Nos decidimos, además, por no incluir tampoco las pocas pruebas no paramétricas. Aunque en algunos estudios de potencia se propone como solución para introducir las pruebas no paramétricas la conversión a sus análogas paramétricas (Cohen, 1962; Sedlmeier y Gigerenzer, 1989), ha quedado demostrado que este procedimiento conlleva sobreestimaciones de la potencia (Rossi, 1990).

Al considerar las pruebas estadísticas, distinguimos entre pruebas centrales, periféricas y de diagnóstico. Como pruebas centrales entendemos todas aquellas sobre las que se focaliza principalmente la investigación; es por ello que, en la mayoría de los casos, para detectarlas hay que tener en cuenta las hipótesis fundamentales sobre las que trabaja el autor. Pruebas periféricas son todas aquellas que no constituyen un objetivo último en la investigación, sino que más bien se utilizan como complemento de las principales o se centran en aspectos secundarios. Por último, el objetivo de las pruebas de diagnóstico es comprobar si se cumplen supuestos estadísticos (por ejemplo, prueba de homogeneidad de Bartlett, pruebas de bondad de ajuste, de fiabilidad intercodificadores, etc.). De acuerdo con estudios anteriores similares, realizamos el cálculo de la potencia exclusivamente sobre las pruebas centrales.

Así, las pruebas estadísticas sobre las que podía determinarse la potencia estadística fueron:  $T$  de diferencias entre medias,  $z$  de diferencias entre proporciones,  $T$  de diferencias entre correlaciones de Pearson (u otros coeficientes de correlación),  $z$  de diferencias entre correlaciones, prueba  $T$  para la significación de un coeficiente de correlación,  $F$  de ANOVA,  $F$  de ANCOVA,  $F$  de regresión simple y múltiple, Chi-cuadrado de Pearson (y derivadas) y las pruebas multivariantes MANOVA, MANCOVA y análisis discriminante.

La unidad de análisis elegida fue el artículo. Si un artículo incluía más de un experimento se calculó la potencia estadística y el tamaño muestral necesario para conseguir una potencia deseable de .80 para cada experimento por separado y después se promediaron, obteniéndose así para todos los experimentos de cada uno de esos estudios valores únicos de potencia y de tamaño muestral.

Para cada experimento seleccionado como unidad de análisis, y manteniendo constante un nivel de significación  $\alpha$  (bilateral) de .05, se calcularon: (a) la potencia estadística para los tamaños del efecto convencionales alto, medio y bajo correspondientes a cada prueba (siguiendo los propuestos por Cohen, 1988); (b) el TE estimado obtenido en el estudio; (c) la potencia estadística correspondiente

al TE estimado; (d) el tamaño muestral que hubiera sido necesario para conseguir una potencia del 80%, tomando el TE estimado para dicho cálculo.

Para la determinación de la potencia estadística y del tamaño muestral necesario para una potencia de .80, nos servimos de las tablas confeccionadas por Cohen (1988) y del programa de computador del que él mismo es co-autor (Borenstein y Cohen, 1988). Sin embargo, para aquellos casos en los que el cálculo de la potencia implicaba una prueba F no central nos resultó más útil el programa de computador STATPOWER de Bavry (1991).

## Resultados

### Características descriptivas

El examen de los trabajos publicados en *Anales de Psicología* pone de manifiesto, en primer lugar, una ligera tendencia hacia los estudios de corte teórico frente a los empíricos, ya que de las 79 publicaciones a lo largo de su historia, 44 de ellas (62%) son teóricas, siendo las 30 restantes (38%) estudios primarios. Entre los estudios teóricos, el 89'8% son revisiones o reflexiones teóricas (44 artículos).

Tabla 2: Relación de estudios incluidos y excluidos según el diseño

Tipo de diseño	Selección		Totales
	Incluidos	Excluidos	
Experimental	4	2 <sup>a</sup>	6
Cuasi-experimental	9	4 <sup>b</sup>	13
Correlacional/Encuesta	3	3 <sup>c</sup>	6
Diseño de caso único	0	5	5
Totales	16	14	30

<sup>a</sup> Uno de los estudios se excluyó por falta de datos; el otro por ser descriptivo.

<sup>b</sup> Dos de los estudios se excluyeron por falta de datos; los otros dos por utilizar pruebas estadísticas para las que no se puede calcular la potencia.

<sup>c</sup> Uno de los estudios se excluyó por falta de datos; los otros dos por ser descriptivos.

En cuanto a los estudios empíricos (véase Tabla 2), de los 30 nos encontramos con cinco diseños de caso único, tres estudios descriptivos y dos estudios que aplicaron pruebas estadísticas para las que no es posible determinar la potencia. Los 20 estudios restantes componían, pues, la población de investigaciones de nuestro estudio de la potencia; no obstante, hubo que desechar cuatro de ellas por no aportar datos suficientes para la realización de los cálculos, quedando en 16 el conjunto de estudios analizados.

En las 16 investigaciones analizadas, predominan los estudios de corte cuasi-experimental (56%), seguidos de los estudios experimentales (25%) y de los correlacionales/encuesta (18'7%). Los cálculos de potencia se han basado en un total de 100 contrastes de hipótesis repartidos tal y como se muestra en la Tabla 3. Predominan las pruebas *T* de Student de diferencias entre medias (35%), tanto para muestras dependientes como independientes, las pruebas *F* de ANOVA (27%) y las pruebas *z* de diferencias entre proporciones (19%). Debe tenerse en cuenta, no obstante, que estos porcentajes no tienen por qué ser necesariamente representativos de la investigación publicada en la revista, ya que en la clasificación de la Tabla 3 no se incluyen las pruebas estadísticas para las que no se calculó la potencia.

Tabla 3: Pruebas estadísticas utilizadas en los 16 estudios

Prueba estadística	Frecuencia
<i>t</i> de Student	35
<i>F</i> de ANOVA	27
<i>z</i> : diferencias entre proporciones	19
Chi-Cuadrado Pearson	10
<i>F</i> de regresión múltiple	8
MANOVA	1
Total	100

En cuanto a la procedencia de los estudios empíricos, se pone de manifiesto el carácter endogámico de la revista, ya que 19 de los 30 estudios empíricos (el 63'3%) proceden de la provincia de Murcia, repartiéndose el resto entre Madrid (seis estudios), Valencia (tres estudios), Barcelona (un estudio) y La Coruña (un estudio). Esta tendencia se acrecienta en los 16 estudios analizados, ya que 12 de ellos procedían de Murcia (75%), siendo los demás de Madrid (un estudio), Valencia (dos estudios) y Barcelona (un estudio).

### Potencia estimada

Los resultados del estudio de potencia se presentan en la Tabla 4 y en el gráfico de cajas múltiples (*multiple boxplot*) expuesto en la Figura 1. Una primera inspección de la Tabla parece poner de manifiesto que la potencia estadística de la investigación publicada en *Anales de Psicología* no es muy diferente de la encontrada en los estudios de potencia publicados en revistas de ámbito internacional. La potencia media alcanzada para un TE bajo es de .13, de .47 para TEs medios y de .76 para TEs altos. Sin embargo, la potencia media calculada a partir de los TEs estimados de cada estudio se eleva a .71, un valor próximo al mínimo (.80) recomendado por Cohen (1988). Es más, si tomamos en consideración el intervalo de confianza (al 95%) obtenido a partir de la potencia media estimada (.71), se observa que la potencia de los estudios analizados no difiere significativamente del valor .80 (intervalo de confianza: .607 - .813). Este patrón de resultados se repitió básicamente para los tres tipos de diseño (experimental, cuasi-experimental y correlacional/encuesta).

Tabla 4: Potencia de los 16 estudios publicados en *Anales de Psicología*

	Tamaños del Efecto			Estimado
	Pequeño	Medio	Alto	
N	16	16	16	16
Mínimo	.06	.16	.41	.36
Máximo	.44	.99	.99	.99
Media	.13	.47	.76	.71
Mediana	.09	.38	.76	.75
Desviación típica	.11	.26	.20	.21
Q <sub>1</sub> (cuartil 1)	.08	.29	.63	.51
Q <sub>3</sub> (cuartil 3)	.14	.68	.95	.86
Intervalo confianza 95%	.077-.182	.341-.599	.663-.857	.607-.813

Un análisis individualizado de la potencia alcanzada por cada estudio refleja, no obstante, la escasa atención que habitualmente se concede a la potencia, ya que de los 16 estudios analizados ninguno alcanzó el criterio de potencia mínima de .80 recomendado por Cohen (1977) para TEs pequeños; tan sólo dos lo alcanzaron para TEs medios (12'5%), siete para TEs altos (43'7%) y seis para los TEs estimados (37'5%). Y, por supuesto, ningún estudio comentó ni efectuó un análisis previo de la potencia de su investigación.

En lo que respecta a la evolución de la potencia estadística en la revista a través de sus ocho años de existencia, en la Tabla 5 puede apreciarse la ausencia de una tendencia clara, sino más bien un patrón errático, tanto para TEs pequeños como medios y altos. Para los TEs estimados se observa, no obstante, una cierta mejora de la potencia estadística con el tiempo, pero dicha apreciación no es corroborada estadísticamente, [ $F(1, 14) = 2'174, p = .162, R_{aj}^2 = .073$ ].

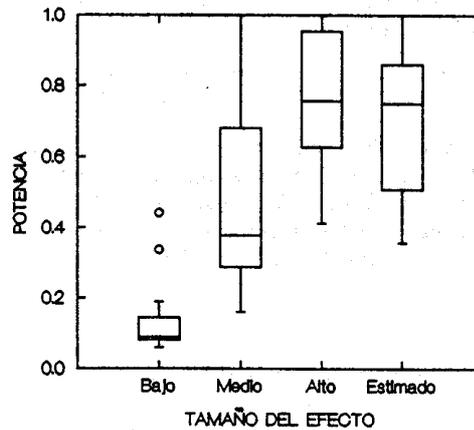


Figura 1.

Tabla 5: Evolución de la potencia estadística de *Anales de Psicología*.

Año	N	Tamaños del Efecto			
		Pequeño	Medio	Alto	Estimado
1984	2	.06	.18	.46	.59
1985	3	.09	.43	.79	.57
1986	2	.10	.38	.74	.76
1987	4	.11	.42	.71	.74
1989	1	.34	.93	.99	.81
1990	3	.23	.70	.89	.74
1991	1	.09	.46	.88	.90

Nota.- En 1988 la revista no publicó ningún número.

### Comparación con otros estudios de potencia

Con objeto de comparar los resultados obtenidos en *Anales de Psicología* con los encontrados en los diversos estudios de potencia recogidos en la Tabla 1, y siguiendo a Rossi (1990) hemos aplicado pruebas  $t$  de una muestra, tomando como parámetro de comparación la potencia media presentada en la Tabla 1 (para TEs pequeño, medio y alto), y como estadístico la potencia media obtenida en nuestro estudio (para los mismos TEs). Los resultados ponen de manifiesto que nuestros niveles de potencia se encuentran significativamente por debajo de los normativos, al menos para TEs pequeño [ $t(15) = 3'521, p(\text{bilat.}) = .004, w^2 = .416$ ] y medio [ $t(15) = 2'234, p(\text{bilat.}) = .042, w^2 = .199$ ], y no así para TEs altos [ $t(15) = 1'743, p(\text{bilat.}) = .101, w^2 = .113$ ].

Al hacer esta misma comparación, Rossi (1990) obtuvo resultados inferiores a los nuestros, ya que incluso con TEs altos la potencia media se encontraba significativamente por debajo del conjunto normativo.

Estos resultados deben ser ponderados teniendo en cuenta la procedencia de las revistas incluídas en la Tabla 1 y que han sido consideradas como grupo normativo de comparación. En dicha tabla se incluyen no sólo estudios de potencia de la investigación psicológica, sino también de otras disciplinas más o menos afines, tales como educación, sociología o economía. Por tanto, surge la necesidad de contrastar nuestros resultados con los estudios de potencia específicos en el ámbito de la psicología.

Además del estudio original de Cohen (1962), los únicos estudios de potencia sobre la investigación psicológica hasta la fecha realizados son los de Chase y Chase (1976), Sedlmeier y Gigerenzer (1989) y Rossi (1990). Sedlmeier y Gigerenzer (1989) realizaron sus cálculos de potencia tomando los criterios originales de TE propuestos por Cohen (1962), no siendo directamente comparables con nuestros cálculos ni con los de Chase y Chase (1976), basados ambos en los criterios de TE más recientes planteados por el propio Cohen (1969, 1977). Nuestros resultados (véase Tabla 4) no parecen apartarse mucho de los obtenidos por Chase y Chase (1976) para TEs pequeño (.25), medio (.67) y grande (.86), si bien algo inferiores. Esto podría explicarse por el elevado tamaño muestral de los estudios (tamaño muestral medio = 350) de la revista analizada por Chase y Chase (1976), ligeramente superior al tamaño encontrado por nosotros (tamaño muestral medio = 99,5; véase Tabla 6).

Pero a nuestro entender, la comparación más interesante para valorar la potencia estadística de la revista *Anales de Psicología* es utilizar el estudio más reciente de Rossi (1990), en el cual analiza la potencia de los artículos publicados en 1982 en las prestigiosas revistas *Journal of Abnormal Psychology*, *Journal of Consulting and Clinical Psychology* y *Journal of Personality and Social Psychology*. Para ello, hemos aplicado pruebas *t* de diferencias entre medias independientes para cada uno de los tres TEs. Aunque la potencia media para los tres TEs de nuestro estudio resulta algo inferior a la obtenida por Rossi (1990), las diferencias encontradas no son significativas (TE bajo:  $t(235) = 1'125, p(\text{bilat.}) = .262, w^2 = .001$ ; TE medio:  $t(235) = 1'541, p = .124, w^2 = .006$ ; TE alto:  $t(235) = 1'491, p = .138, w^2 = .005$ ). Por tanto, podemos considerar que la potencia estadística de la investigación psicológica publicada en *Anales de Psicología* es similar a la publicada en revistas de corte internacional.

### Tamaños muestrales

Como complemento al estudio de la potencia, registramos los tamaños muestrales utilizados en las investigaciones publicadas con objeto de explorar esta importante característica del diseño de una investigación. En la primera columna de la Tabla 6 se presentan los estadísticos descriptivos básicos calculados sobre el tamaño muestral total de cada estudio. El primer dato relevante que surge de su inspección es la fuerte asimetría positiva en la distribución de los tamaños muestrales, lo que hace sospechosa la interpretación de la media; resulta, pues, más representativo el tamaño muestral mediano, el cual alcanzó un valor en torno a los 43 sujetos. Este valor no discrepa mucho de los tamaños muestrales medios de 50 ó 60 obtenidos en otros estudios (Holmes, 1979).

**Tabla 6: Tamaño muestral real y estimado para una potencia de .80**

	N(real)	N(.80)	Diferencia
N	16	16	16
Mínimo	15	12	-1322
Máximo	551	1581	78
Media	99'5	248'3	-148'81
Mediana	43'5	93	-31'5
Desviación típica	135'29	418'31	346'08
Q1 (cuartil 1)	30	51	-83
Q3 (cuartil 3)	105	148'5	1

N(real) presenta los tamaños muestrales globales utilizados en los 16 estudios.

N(.80) representa los tamaños muestrales que deberían haberse utilizado para asegurar una potencia de .80.

La columna "Diferencia" es el resultado de restar N(real) - N(.80) para cada estudio.

El estudio de Kazdin y Bass (1989) sobre las investigaciones comparativas de la eficacia de la psicoterapia, también puede servir, salvando las distancias, como punto de referencia. Estos autores encontraron en dicho campo de investigación tamaños muestrales globales en torno a 34 sujetos (para diseños de dos grupos). Así pues, los tamaños muestrales habitualmente utilizados en las investigaciones publicadas en *Anales de Psicología* parecen encontrarse en la línea de los estudios publicados en las revistas psicológicas internacionales. No obstante, es preciso tener en cuenta que la selección del tamaño muestral varía enormemente en función del campo específico y del diseño de la investigación; en consecuencia, generalizaciones de estos resultados deben hacerse con extrema precaución.

#### **Tamaños muestrales "ideales" para la investigación psicológica**

Un procedimiento más adecuado para determinar la adecuación de los tamaños muestrales utilizados en las investigaciones consiste en comparar éstos con los que deberían haberse empleado suponiendo que pretendiéramos garantizar un determinado nivel mínimo de potencia estadística. A excepción de Kazdin y Bass (1989), estos cálculos no han sido efectuados en anteriores estudios de potencia, por lo que no disponemos de puntos de referencia.

Siguiendo las recomendaciones de Cohen (1988), para cada estudio calculamos el tamaño muestral "ideal" fijando una potencia  $1 - \beta = .80$ , manteniendo la tasa de error Tipo I en  $\alpha = .05$  (bilateral) y estimando el TE de cada estudio a partir de los propios resultados estadísticos reportados. Los resultados de estos cálculos, como puede observarse en la segunda columna de la Tabla 6, arrojan un tamaño muestral "ideal" mediano de 93 sujetos, es decir, casi dos veces superior en promedio al realmente utilizado. En la tercera columna se presentan los resultados de calcular, para cada investigación, la diferencia entre el tamaño muestral real y el "ideal". Si bien existe una alta heterogeneidad entre los estudios, hablando en términos de la mediana hubieran sido necesarios unos 31 sujetos más por estudio para asegurar una potencia del 80%. De hecho, sólo cuatro de los 16 estudios (el 25%) utilizó tamaños muestrales superiores al mínimo requerido para asegurar dicha potencia. Esta deficiencia coincide con los resultados obtenidos por Kazdin y Bass (1989).

#### **Discusión**

Con el presente estudio hemos pretendido llamar la atención de los investigadores en ciencias del comportamiento hacia la necesidad de controlar la potencia estadística cuando se realizan investigaciones en las que es preciso aplicar pruebas de significación. Con esta finalidad hemos sometido a los estudios empíricos publicados en *Anales de Psicología* a un análisis de la potencia.

Los resultados obtenidos nos permiten plantear varias conclusiones. En primer lugar, la potencia estadística de la investigación publicada en *Anales de Psicología* no difiere de la encontrada en las más

recientes revisiones de la investigación psicológica publicada en revistas de prestigio internacional. En efecto, nuestros resultados coinciden, en términos generales, con los obtenidos por Rossi (1990) al analizar la potencia exhibida en 1982 por las conocidas revistas *Journal of Abnormal Psychology*, *Journal of Consulting and Clinical Psychology* y *Journal of Personality and Social Psychology*; así como con los de Sedlmeier y Gigerenzer (1989) al revisar el *Journal of Abnormal Psychology* del año 1984; o con los del año 1974 del *Journal of Applied Psychology* (Chase y Chase, 1976).

En segundo lugar, esta conclusión no debe tomarse como motivo de satisfacción, sino más bien como evidencia del lamentable estado de salud en que se encuentra la investigación psicológica, al menos en lo que respecta a la potencia estadística. Este pesimismo cobra mayor fuerza si tenemos en cuenta que, treinta años después de que Cohen (1962) denunciara por primera vez el deficiente estado de la potencia estadística en la investigación psicológica, el panorama internacional no ha cambiado en absoluto. *Anales de Psicología* es un mero reflejo de esta situación.

Esta segunda conclusión nos induce a plantear otra lectura. Obviamente, los criterios de selección de *Anales de Psicología* no son tan estrictos como puedan serlo los de las revistas arriba mencionadas. Por tanto, el hecho de que la potencia exhibida por las publicaciones de éstas y aquélla sea similar parece evidenciar la escasa o nula atención que los consejos editoriales y/o los revisores dedican a este aspecto tan crucial del análisis estadístico.

### ¿Es necesaria la potencia estadística?

Este absoluto desinterés manifestado por los investigadores hacia la potencia nos induce a preguntarnos si realmente necesitamos controlar la potencia estadística para hacer una investigación. La respuesta es afirmativa. Cuando realizamos una investigación en condiciones de baja potencia, la aceptación de la  $H_0$  no nos permitirá llegar a ninguna conclusión fiable, ya que la probabilidad de cometer un error tipo II es alta. Pero si, además, estamos estudiando un TE pequeño, no podremos estar seguros de nuestra decisión ni siquiera cuando rechazamos la  $H_0$ , ya que debido al conocido sesgo de publicación en favor de los resultados significativos (Bakan, 1966; Cohen, 1965; Coursol y Wagner, 1986) es muy probable que cometamos un error tipo I, aun cuando lo hayamos fijado nominalmente en un nivel bajo (por ej., .05). Por otra parte, sea cual sea la magnitud del efecto que estemos estudiando, trabajar con baja potencia limita las posibilidades de replicación del estudio, provocando el que en un mismo tema de investigación se obtengan resultados significativos y no significativos.

La necesidad de conocer a priori la potencia estadística es menos imperiosa cuando estamos estudiando un TE de magnitud elevada, aunque incluso en esta situación tendríamos problemas de replicación. Sin embargo, es un hecho demostrado (Cooper y Findley, 1982; Haase, Waechter y Solomon, 1982) que la investigación psicológica se mueve dentro del rango de magnitudes del efecto consideradas como medias, según la clasificación de Cohen (1988). La gran proliferación de estudios meta-analíticos en los últimos quince años parece confirmar las intuiciones de Cohen.

Como consecuencia de lo dicho, la siguiente pregunta es obligada: ¿por qué no dedican los investigadores psicológicos la atención que merece la potencia? La razón última hay que buscarla en las raíces de la teoría del contraste de hipótesis que domina la investigación en ciencias sociales y del comportamiento. Esta no es otra que el enfoque Fisheriano del contraste de hipótesis, el cual se basa en el planteamiento únicamente de hipótesis nulas, imposibilitando la determinación de la potencia del contraste (Gigerenzer, 1987; Gigerenzer y Murray, 1987). No fue sino después de la IIª Guerra Mundial cuando se incorporó, de una manera forzada, la teoría de Neyman y E.S. Pearson del contraste de hipótesis, que abogaba en favor del planteamiento de hipótesis alternativas y del control de la potencia. Los textos de estadística aplicada a la investigación social y psicológica con los que nos hemos formado presentan una teoría híbrida entre ambos enfoques, no exenta de contradicciones y con un escaso hincapié en el concepto de potencia (Gigerenzer, Swijtink et al., 1989). Así pues, "... el constante olvido de la potencia parece ser una consecuencia directa de este estado de la cuestión" (Sedlmeier y Gigerenzer, 1989, p. 314).

Las dificultades que entraña al cálculo de la potencia estadística también pueden haber contribuido a su escaso uso. Afortunadamente, el panorama está cambiando gracias a la publicación de programas de software elaborados específicamente para facilitar la realización de estos cálculos (Bavry, 1991; Borenstein y Cohen, 1988; Dallal, 1988), así como de manuales sobre potencia (Cohen, 1988; Kraemer y Thiemann, 1987; Lipsey, 1990).

Otro factor implicado en la falta de atención hacia la potencia es su dependencia de la magnitud del efecto que se pretende probar. En ausencia de estimaciones del TE, los cálculos de potencia sólo pueden ir guiados por las definiciones normativas de TE pequeño, medio y grande propuestas por Cohen (1969). Sin embargo, el creciente número de estudios meta-analíticos en todos los campos de la investigación psicológica, con su énfasis en la magnitud del efecto, está facilitando la incorporación rutinaria de los cálculos de la potencia en la investigación. En efecto, en los meta-análisis el investigador dispone de estimaciones precisas del TE que pretende estudiar, evitando tener que recurrir a los criterios normativos del TE en sus cálculos. Pero el principal impulso que, sin duda, puede recibir la potencia procederá de los propios consejos editoriales cuando exijan el control de este parámetro como requisito para la publicación de investigaciones.

## Referencias

- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437.
- Bavry, J.L. (1991). *STAT-POWER: Statistical Design Analysis System* (2nd ed.). Chicago, IL: Scientific Software, Inc.
- Borenstein, M. y Cohen, J. (1988). *Statistical Power Analysis: A Computer Program*. Hillsdale, NJ: Erlbaum.
- Brewer, J.K. (1972). On the power of statistical tests in the American Educational Research Journal. *American Educational Research Journal*, 9, 391-401.
- Brewer, J.K. y Owen, P.W. (1973). A note on the power of statistical tests in the Journal of Educational Measurement. *Journal of Educational Measurement*, 10, 71-74.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal & Social Psychology*, 65, 145-153.
- Cohen, J. (1965). Some statistical problems in psychological research. En B.B. Wolman (Ed.), *Handbook of Clinical Psychology* (pp. 95-121). New York: McGraw-Hill.
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (rev. ed.). New York: Academic Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Academic Press.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.
- Cooper, H.M. y Findley, M. (1982). Expected effect sizes: Estimates for statistical power analysis in social psychology. *Personality & Social Psychology Bulletin*, 8, 168-173.
- Coursol, A. y Wagner, E.E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology: Research & Practice*, 17, 136-137.
- Chase, L.J. y Baran, S.J. (1976). An assessment of quantitative research in mass communication. *Journalism Quarterly*, 53, 308-311.
- Chase, L.J. y Chase, R.B. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology*, 61, 234-237.
- Chase, L.J.; Chase, R.B. y Tucker, R.K. (1978). Statistical power in physical anthropology: A technical report. *American Journal of Physical Anthropology*, 49, 133-138.
- Chase, L.J. y Tucker, R.K. (1975). A power-analytic examination of contemporary communication research. *Speech Monographs*, 42, 29-41.
- Christensen, J.E. y Christensen, C.E. (1977). Statistical power analysis of health, physical education, and recreation research. *Research Quarterly*, 48, 204-208.
- Daly, J.A. y Hexamer, A. (1983). Statistical power in research in English education. *Research in the Teaching of English*, 17, 157-164.
- Dallal, G.E. (1988). *DESIGN: A Supplementary Module for SYSTAT* (Vers. 2.0). Evanston, IL: SYSTAT, Inc.

- Gigerenzer, G. (1987). Probabilistic thinking and the fight against subjectivity. En L. Krüger, G. Gigerenzer y M. Morgan (Eds.), *The Probabilistic Revolution: Vol. 2. Ideas in the Sciences* (pp. 11-33). Cambridge, MA: MIT Press.
- Gigerenzer, G. y Murray, D.J. (1987). *Cognition as Intuitive Statistics*. Hillsdale, NJ: Erlbaum.
- Gigerenzer, G.; Swijtink, Z.; Porter, T.; Daston, L.J.; Beatty, J. y Krüger, L. (1989). *The Empire of Chance: How Probability Changed Science and Everyday Life*. Cambridge, England: Cambridge University Press.
- Haase, R.F. (1974). Power analysis of research in counselor education. *Counselor Education & Supervision*, 14, 124-132.
- Haase, R.F.; Waechter, D.M. y Solomon, G.S. (1982). How significant is a significant difference? Average effect size of research in Counseling Psychology. *Journal of Counseling Psychology*, 29, 58-65.
- Hays, W. (1988). *Statistics* (4th ed.). New York: Holt, Rinehart & Winston.
- Holmes, C.B. (1979). Sample size in psychological research. *Perceptual & Motor Skills*, 49, 283-288.
- Jones, B.J. y Brewer, J.K. (1972). An analysis of the power of statistical tests reported in the Research Quarterly. *Research Quarterly*, 43, 23-30.
- Katzer, J. y Sodt, J. (1973). An analysis of the use of statistical testing in communication research. *Journal of Communication*, 23, 251-265.
- Kazdin, A.E. y Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting & Clinical Psychology*, 57, 138-147.
- Kraemer, H.C. y Thiemann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Newbury Park, CA: Sage.
- Kroll, R.M. y Chase, L.J. (1975). Communication disorders: A power analytic assessment of recent research. *Journal of Communication Disorders*, 8, 237-247.
- Levenson, R.L. (1980). Statistical power analysis: Implications for researchers, planners, and practitioners in gerontology. *Gerontologist*, 20, 494-498.
- Lipsey, M.W. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.
- Mazen, A.M.; Graf, L.A.; Kellog, C.E. y Hemmasi, M. (1987). Statistical power in contemporary management research. *Academy of Management Journal*, 30, 369-380.
- Mazen, A.M.; Hemmasi, M. y Lewis, M.F. (1987). Assessment of statistical power in contemporary strategy research. *Strategic Management Journal*, 8, 403-410.
- Orme, J.G. y Combs-Orme, T. (1986). Statistical power and type II errors in social work research. *Social Work Research & Abstracts*, 22, 3-10.
- Orme, J.G. y Tolman, R.M. (1986). The statistical power of a decade of social work education research. *Social Service Review*, 60, 619-632.
- Ottenbacher, K. (1982). Statistical power and research in occupational therapy. *Occupational Therapy Journal of Research*, 2, 13-25.
- Pennik, J.E. y Brewer, J.K. (1972). The power of statistical tests in science teaching research. *Journal of Research in Science Teaching*, 9, 377-381.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1990). Replication in behavioral research. *Journal of Social Behavior & Personality*, 5, 1-30.
- Rosnow, R.L. y Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology*, 35, 203-208.
- Rossi, J.S. (1983, abril). *Inadequate statistical power: A source of artifactual controversy in behavioral research*. Paper presented at the 54th annual meeting of the Eastern Psychological Association, Philadelphia, PA.
- Rossi, J.S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting & Clinical Psychology*, 58, 646-656.
- Sawyer, A.G. y Ball, A.D. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research*, 18, 275-290.
- Sedlmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309-316.
- Spreitzer, E. y Chase, L.J. (1974). *Statistical power in sociological research: An examination of data-analytic strategies*. Unpublished manuscript, Bowling Green State University, Department of Sociology, Bowling Green, OH. (Citado en Rossi, 1990, p. 648.)

Woolley, T.W. (1983). A comprehensive power-analytic investigation of research in medical education. *Journal of Medical Education*, 58, 710-715.

Woolley, T.W. y Dawson, G.O. (1983). A follow-up power analysis of the statistical tests used in the Journal of Research in Science Teaching. *Journal of Research in Science Teaching*, 20, 673-681.

*(Original recibido: 11-6-1992)*

*(Original aceptado: 14-7-1992)*