

## Consideraciones generales en torno a los estudios de potencia

Roser Bono\* y Jaume Arnau

Universidad de Barcelona

**Resumen:** Debido a la creciente importancia del análisis de la potencia en ciencias del comportamiento y a una escasa y ambigua literatura durante décadas, el objetivo de este trabajo es presentar el marco conceptual básico del análisis estadístico de la potencia. Para ello, se lleva a cabo una revisión histórica del tema, y se discute lo más relevante acerca de aquellos elementos que afectan directamente a la potencia: tamaño muestral, nivel de significación y tamaño del efecto.

**Palabras Clave:** Ciencias conductuales, Inferencia estadística, Potencia estadística, Tamaño muestral, Nivel de significación, Tamaño del efecto.

**Title:** General considerations about statistical power studies.

**Abstract:** Due to the increasing importance of power analysis in behavioral sciences and to the fact that the available literature on the subject has been scarce and ambiguous for decades, the objective of this study is to provide the basic conceptual framework of statistical analysis of power. For this purpose, an historical review is made of the subject, and the most relevant aspects of the elements which directly influence power are discussed: sample size, significance level and effect size.

**Key Words:** behavioral sciences, statistical inference, statistical power, sample size, significance level, effect size.

### Desarrollo histórico

El concepto de potencia se atribuye, históricamente, a Neyman y Pearson (1928, 1933). A partir de entonces, aparecieron una serie de autores que tenían en cuenta en sus estudios la potencia estadística (Cox, 1948; McNemar, 1960; Sterling, 1959; Tukey, 1960; Tullock, 1959). Dentro de la literatura de las ciencias sociales, es de destacar Mosteller y Bush (1954). Sin embargo, no fue hasta 1962 cuando apareció por primera vez, en el ámbito de las ciencias sociales, un estudio sistemático de la potencia estadística. Se trata del trabajo realizado por Cohen (1962) en el que se destacó, en primer lugar, la importancia de la potencia estadística dentro de la investigación experimental y, en segundo lugar, proporcionó una serie de pautas para llevar a cabo un análisis de potencia. Este estudio alentó a los investigadores a prestar mayor atención a la potencia de las pruebas y a no centrarse, exclusivamente, en el nivel de significación. Posteriormente, Cohen (1965, 1969) publicó

una serie de recomendaciones para tratar con eficacia el problema de la potencia y, más recientemente, elaboró unas tablas muy útiles para su cálculo (Cohen, 1988). Sin duda, el trabajo de Cohen (1988) ha inspirado varios estudios sobre la potencia y el tamaño del efecto en áreas de las ciencias sociales, así como diversos programas de ordenador (examinados en Goldstein, 1989). Otros investigadores, que también promovieron el interés por el tema, fueron los sociólogos Labovitz (1968) y Skipper, Guenther y Nass (1967) y los psicólogos Bakan (1966), Lykken (1968) y Runyon (1969). Sin embargo, esta proliferación de estudios no ha tenido un impacto en investigaciones posteriores. Así, de 54 artículos publicados en el volumen de 1984 del *Journal of Abnormal Psychology*, sólo dos mencionaban la potencia y en ninguno de ellos se estimaba (Sedlmeier y Gigerenzer, 1989). Este escaso interés de los investigadores por el control de la potencia estadística queda reflejado también en una serie de estudios de potencia realizados en base a diversas revistas españolas, como son *Anales de Psicología* (Sánchez *et al.*, 1992), *Revista de Psicología General y Aplicada* (Valera *et al.*, 1993), *Psicológica* (Frías, García y Pascual, 1993) y *Anuario de Psicología* (Pascual, Frías y García, 1993).

\* **Dirección para correspondencia:** Roser Bono. Departamento de Metodología de las Ciencias del Comportamiento. Facultad de Psicología, Universidad de Barcelona. Passeig de la Vall d'Hebron, 171. 08035-Barcelona (España).  
E-mail: rbono@psi.ub.es

© Copyright 1995: Servicio de Publicaciones de la Universidad de Murcia, Murcia (España). ISSN: 0212-9728. Artículo recibido: 30-5-95, aceptado: 28-7-95.

Además, si se toman como evidencia manuales de estadística, se observa que con frecuencia la potencia no es tratada.

No se sabe con certeza la razón por la que los investigadores dejan de lado el análisis de la potencia. Según Cohen (1992), una buena parte se debe a la poca importancia que se da al tamaño del efecto y a la mayor relevancia de los resultados estadísticos y los valores de  $\alpha$ . Chase y Tucker (1976) y Sedlmeier y Gigerenzer (1989) atribuyen este hecho al predominio de la teoría fisheriana. La escuela de Fisher considera las pruebas estadísticas como pruebas de significación (Fisher, 1935, 1950, 1955), mientras que la escuela de Neyman-Pearson las conceptúa como pruebas de decisión (Neyman y Pearson, 1928, 1933, 1936).

Según Chase y Tucker (1976), esta controversia entre tradiciones puede reducirse a que ante una diferencia significativa, un defensor del enfoque fisheriano concluiría que la hipótesis nula ( $H_0$ ) no es válida (con lo que se prueba la existencia del fenómeno que se esté estudiando); mientras que un investigador de la tradición de Neyman-Pearson, simplemente rechazaría la  $H_0$  en esta ocasión particular.

Otra distinción importante entre estas dos aproximaciones se refiere a si el nivel de significación empleado en el análisis debe fijarse anticipadamente. Aunque Fisher (1935, 1950, 1955) daba prioridad a un nivel de significación de 0.05, nunca prescribió que tal nivel debiera mantenerse fijo o que debiera establecerse antes de llevar a cabo el experimento. Por contra, la posición de Neyman-Pearson (1928, 1933, 1936) requiere que el nivel de significación se determine antes de cualquier análisis estadístico y que el investigador se adhiera a él en todas las decisiones estadísticas. En este sentido, si el nivel de significación no se estableciera con anterioridad, se podría caer en el error de fijar un valor a basarse más en los resultados que en una estimación de la probabilidad.

Una tercera divergencia entre las dos teorías se refiere a la interpretación de resultados positivos y negativos de la investigación experimental. De acuerdo con Fisher (1935,

1950, 1955), se puede afirmar que el efecto no es cero cuando se rechaza la  $H_0$ , pero no es posible concluir que sea cero cuando se acepta. Lo que realmente significa la afirmación falsa de que la  $H_0$  es cierta es que existe un efecto despreciable o trivial (Cohen, 1988). Por el contrario, el planteamiento de Neyman-Pearson (1928, 1933, 1936) postula la existencia de una hipótesis alternativa ( $H_1$ ) exacta del tamaño del efecto. Esta proposición llevó a Neyman y Pearson (1928, 1933) al concepto de error de Tipo II (aceptar la  $H_0$  cuando es falsa), relacionado con el de potencia. En cuanto al error de Tipo I, característico de la tradición fisheriana, cabe destacar que no consiste en rechazar la  $H_0$  cuando es verdadera sino en el nivel de riesgo que el investigador está dispuesto a aceptar cuando interpreta los resultados de una prueba estadística.

En conclusión, la razón de porque los psicólogos descuidan la potencia se debe a razones históricas. Los manuales de psicología y educación transmitieron, primero, el mensaje fisheriano y después de la II Guerra Mundial, los autores comenzaron a darse cuenta del impacto de la aproximación de Neyman-Pearson y a sustituir la teoría anterior. Sin embargo, los conceptos de error Tipo II y potencia fueron incluidos, por la mayoría de investigadores, dentro del marco fisheriano. El resultado fue una teoría híbrida con ideas contrapuestas (prueba de la  $H_0$ , siguiendo a Fisher, y error de Tipo II y especificación del error de Tipo I antes de la recogida de datos, siguiendo a Neyman y Pearson), lo que conllevó una confusión del significado de conceptos básicos (Bakan, 1966; Oakes, 1986) y, en algún grado, una explicación del continuo descuido de la potencia. Esta estadística híbrida se enseña, normalmente, como estadística, sin mención de los originadores de las respectivas ideas (Gigerenzer y Murray, 1987).

Según Sedlmeier y Gigerenzer (1989), este continuado desinterés por la potencia de las pruebas estadísticas sólo cambiará cuando los editores de las principales revistas exijan, dentro de su política editorial, que los autores estimen la potencia de sus pruebas de signifi-

cación. De todos modos, en los últimos años han aparecido una serie de textos básicos sobre potencia estadística (Cohen, 1988; Kraemer y Thiemann, 1987; Lipsey, 1990, entre los más destacados). Desafortunadamente, la mayor parte de estas fuentes aún no contienen mucha información sobre cómo calcular la potencia o están limitadas a la consideración de pruebas tradicionales, tales como *t* y *f*. Por otro lado,

están disponibles programas informáticos como por ejemplo el *POWER* (Borenstein y Cohen, 1988) que calcula la potencia del análisis de la variancia unidireccional (Borenstein *et al.*, 1990), de la regresión múltiple (Rothstein *et al.*, 1990), y de correlaciones y diferencias entre proporciones (Borenstein y Cohen, 1988). Otro programa de mayor complejidad es el *STAT-POWER* (Bavry, 1991).

**Tabla 1.** Posibilidades de error en pruebas de significación estadística. (Tomado de Lipsey, 1990).

Conclusión de la prueba estadística	Situación en la población	
	T y C difieren	T y C no difieren
Diferencias significativas (rechazar $H_0$ )	Conclusión correcta Probabilidad = $1-\beta$ (potencia)	Error de Tipo I Probabilidad = $\alpha$
Diferencias no significativas (aceptar $H_0$ )	Error de Tipo II Probabilidad = $\beta$	Conclusión correcta Probabilidad = $1-\alpha$

(T = Grupo de Tratamiento; C = Grupo Control)

### Concepto de potencia: su relación con errores de Tipo I y Tipo II

Es sabido que al tomar una decisión estadística (rechazar o aceptar la  $H_0$ ) se corre el riesgo de cometer un error, debido a la dificultad de discriminar con certeza entre las hipótesis falsas y las verdaderas. Estos errores son, esencialmente, dos: a) error de Tipo I y b) error de Tipo II (tabla 1). Las dos clases de error tienen diferentes consecuencias y costos, por lo que la probabilidad de cometer ambos errores se intenta mantener tan baja como sea posible.

El error de Tipo I o nivel de significación ( $\alpha$ ) es el más familiar para la mayoría de investigadores y, por lo común, está fijado en 0.05. Al aplicar una prueba estadística también existe la posibilidad de cometer un error de Tipo II, denotado por  $\beta$ , que es la probabilidad de aceptar la  $H_0$  cuando es falsa. Estos dos errores ( $\alpha$  y  $\beta$ ) están inversamente relacionados. En efecto, al controlar el error de Tipo I, el error de Tipo II queda determinado (Cohen, 1988; Rossi, 1990; Stevens, 1986). Así,

dada una muestra, al disminuir a aumenta  $\beta$  y si se desea disminuir ambos errores, la única forma consiste en aumentar el tamaño de muestra, que no siempre es posible. En consecuencia, el problema que surge al investigador es el de alcanzar un equilibrio entre  $\alpha$  y  $\beta$ . Un análisis de los costes, beneficios y otros factores relevantes asociados a los errores Tipo I y II se halla en Brown (1983), Cascio y Zedeck (1983), Nagel y Neef (1977) y Schneider y Darcy (1984).

La probabilidad de un error Tipo I viene designada de antemano por  $\alpha$ , de modo que el riesgo de rechazar  $H_0$  verdaderas queda controlado. Por contra, la tasa de error Tipo II no se puede determinar hasta que se conozca la  $H_1$ , o sea, hasta que el investigador sepa el efecto concreto existente. Por lo tanto, el experimentador deberá basarse en estimaciones.

En la teoría de Neyman-Pearson, la potencia de una prueba estadística es la probabilidad de resultados significativos; es decir, la aceptación de la  $H_1$  cuando es verdadera (Cohen,

1988; Lipsey, 1990). En otras palabras, la potencia de un estadístico viene dada por su capacidad de rechazar la  $H_0$  correctamente, de modo que está determinada por la probabilidad de cometer errores de Tipo II (tabla 1). Así, la potencia de una prueba es el complemento de la probabilidad de un error Tipo II ( $1-\beta$ ). Cuando  $b$  es pequeño, entonces la potencia es grande y viceversa.

Cohen (1992) propone por convención una potencia de 0.80 (o sea,  $\beta=0.20$ ). Un valor sustancialmente inferior a 0.80 implicaría un gran riesgo de incurrir en un error de Tipo II, y un valor superior exigiría una muestra muy grande, fuera de los recursos del investigador. Si se toma un valor  $\alpha$  convencional de 0.05 y una potencia de 0.80, la resultante de  $\beta:\alpha$  es 4:1 (0.20 a 0.05).

Según Strube (1985), el hecho de rechazar la  $H_1$  debido a una potencia baja lleva, en un nivel de meta-análisis, a abandonar prematuramente una área de investigación prometedor. Por ello, dada la influencia que tienen las revisiones de estudios anteriores para una investigación futura, es indispensable que las decisiones concernientes a las hipótesis sean correctas.

Por último, antes de presentar los distintos elementos que determinan la potencia, se detallan a continuación las consecuencias positivas que tiene la consideración de la potencia estadística (Rossi, 1990):

- 1) La potencia de una prueba es la probabilidad de obtener un resultado estadísticamente significativo. Así, si las estimaciones de potencia obtenidas *a priori* son bajas, el investigador puede elegir entre aumentar la potencia o abandonar la investigación, en caso que los costes de incrementar la potencia sean demasiado altos, o si la baja potencia no justifica el tiempo, el esfuerzo y la cantidad de recursos que requiere la investigación que se pretende llevar a cabo.
- 2) Conocer la potencia de una prueba estadística facilita la interpretación de los resultados nulos. La incapacidad de rechazar la  $H_0$  no significa que sea cierta, sino que no existen suficientes evidencias para rechazarla. Si

la potencia es baja, es razonable sugerir, *a priori*, que no hay una razón favorable para rechazar la  $H_0$ . En cambio, si la potencia es alta, el no poder rechazar la  $H_0$  puede considerarse -dentro de unos límites- como una afirmación de dicha hipótesis, puesto que la probabilidad de un error de Tipo II será baja.

3) Cuando la potencia estadística media es baja, se cuestiona la validez de los resultados estadísticamente significativos. En este sentido, una proporción sustancial de resultados significativos publicados pueden ser errores de Tipo I. Por otro lado, cuando la potencia es marginal (aproximadamente 0.50) es muy posible obtener un patrón de resultados inconsistentes, que en algunos casos dará lugar a resultados significativos y en otros no (Kazdin y Bass, 1989).

### Elementos de la potencia estadística

Las estimaciones de la potencia estadística quedan determinadas por tres componentes: a) tamaño muestral, es decir cuanto más grande sea la muestra, mayor será la potencia; b) nivel de significación, dado que al incrementarse la probabilidad de error Tipo I, la potencia también aumenta, y c) tamaño del efecto o grado en que el fenómeno de interés está presente, de modo que cuando los efectos son grandes, la potencia incrementa.

En este apartado presentamos una breve exposición de cada uno de estos elementos (para una explicación más amplia, puede consultarse a Cohen, 1988, 1992). Sin embargo, el tamaño del efecto se describe con más detalle debido a que, por un lado, es una de las formas más importantes para determinar la potencia o el tamaño muestral requerido -o ambos- y, por otro, es el concepto, relacionado con la inferencia estadística, menos familiar a los investigadores de las ciencias del comportamiento.

### Tamaño muestral

La potencia aumenta en función del tamaño de la muestra, manteniéndose constantes los demás elementos (nivel de significación y tamaño del efecto). Ello se debe a que al aumentar la muestra, el error aleatorio de medida es menor y la fiabilidad o precisión de los resultados es mayor y, por consiguiente, la probabilidad de darse el fenómeno bajo prueba también es mayor (Cohen, 1988; Lipsey, 1990). Así, la exclusión de fuentes de variabilidad irrelevantes para el conocimiento del fenómeno que se está estudiando es útil para incrementar la potencia.

Cuando se planifica una investigación, es conveniente conocer, para un determinado error  $\alpha$  y un tamaño del efecto hipotético, la muestra necesaria para obtener la potencia deseada. Es preciso incrementar el tamaño muestral al aumentar la potencia y al disminuir el tamaño del efecto y a (Cohen, 1992).

### Nivel de significación

El nivel de significación es un estadístico de prueba en virtud del cual se concluye acerca de la existencia del fenómeno, o el riesgo de rechazar erróneamente la  $H_0$ . Para hallar la potencia, de acuerdo con la teoría de Neyman-Pearson, el valor de  $\alpha$  debe establecerse por anticipado. A no ser que se indique lo contrario, se toma  $\alpha=0.05$  (es parte del legado fisheriano); aunque evidentemente pueden seleccionarse otros valores.

Asumiendo que un fenómeno concreto existe en un determinado grado en la población y manteniéndose los demás elementos constantes, cuanto más estricto es el estándar de prueba ( $\alpha$ ), menor es la probabilidad de que la muestra lleve a resultados que se ajusten a dicho estándar (Cohen, 1988; Lipsey, 1990; Sedlmeier y Gigerenzer, 1989). Así pues, la práctica de tomar valores  $\alpha$  pequeños -cuanto más pequeños mejor- conduce a una baja potencia. Además, el complemento de la potencia ( $\beta$ ) constituye también un error (el

error de Tipo II), por lo que la inferencia estadística puede ser considerada como una ponderación de los dos tipos de error (Cohen, 1962). Por ejemplo, un investigador puede establecer el riesgo de rechazar equivocadamente la  $H_0$  a un nivel de  $\alpha=0.001$  y, con ello, reducir la potencia de la prueba (por consiguiente, aumentar el valor de  $\beta$ ). Por otro lado, es posible incrementar la potencia de la prueba hasta un valor aceptable usando un nivel de significación menos riguroso. No obstante, de esta forma se incrementa la probabilidad de cometer un error Tipo I. Por lo tanto, debe equilibrarse la probabilidad de incurrir en un error de Tipo I con la probabilidad de incidir en un error de Tipo II, puesto que, como ya se ha comentado anteriormente, no puede reducirse la probabilidad de cometer un tipo de error sin que se incremente el otro.

### Tamaño del efecto

El tamaño del efecto (TE) es el grado en que el fenómeno bajo estudio está presente en la población, o el grado en que la  $H_0$  es falsa (Cohen, 1988), es decir, la discrepancia entre  $H_0$  y  $H_1$ . Así, el TE puede ser tratado como un factor que toma el valor cero cuando la  $H_0$  es verdadera, y cualquier otro valor cuando es falsa. Por lo tanto, se trata de un índice del grado de distanciamiento de la  $H_0$  (Lipsey, 1990). Lo ideal sería que este índice fuera aplicable a todos los ámbitos de investigación; sin embargo, la unidad de medida debe ser apropiada a los datos y a la prueba de significación utilizada. De todas formas, las tablas de potencia existentes para determinadas pruebas tienen un índice libre de unidades, ya que es inviable preparar tablas para cada nueva unidad de medida con la que se trabaje.

Los investigadores del comportamiento sostienen que especificar el TE es la parte más difícil del análisis de potencia. La dificultad se debe, al menos en parte, a la escasa concienciación sobre la magnitud de los fenómenos y a que la teoría fisheriana considera el TE como cualquier valor distinto de cero. Una solución

al respecto es recurrir a investigaciones anteriores dentro del área de estudio. Si esto no es suficiente, un recurso adicional consiste en especificar los TE en términos de "pequeño", "mediano" y "grande" (Cohen, 1988). Se trata, al igual que el nivel de significación de 0.05, de una convención que puede ser mal utilizada si se abusa de esa convencionalidad. Aunque sus definiciones son arbitrarias, deben ser razonables (Cohen, 1988). Por ejemplo, los TE pequeños han de ser lo suficientemente grandes para que su búsqueda no sea una tarea sin sentido; pero no tan grandes como para que sean fácilmente perceptibles. Cohen (1988, 1992) establece los valores convencionales pequeños, medianos y grandes para varias pruebas estadísticas distintas (prueba *t*, análisis de la variancia unidireccional, pruebas de correlación, diferencias entre correlaciones, prueba de los signos, diferencias entre proporciones, prueba  $\chi^2$ , y correlación parcial y múltiple), de manera que los TE medios representan un efecto que el observador cuidadoso puede observar directamente, los TE pequeños son bastante más pequeños que los medios, pero no tanto como para ser triviales, y los TE grandes están a la misma distancia por encima de los medios que los pequeños lo están por debajo.

El TE es difícil de estimar, puesto que es un tema relativamente nuevo en ciencias conductuales. Así, para Cohen (1988), determinar los TE como pequeños, medianos y grandes es el primer paso tentativo en esta dirección. Estas convenciones han quedado fijadas y su uso se ha extendido también al meta-análisis. No obstante, aun así, existe un gran riesgo de malas interpretaciones.

El TE es un elemento crucial en el análisis de potencia, ya que la potencia depende de la magnitud del efecto que se quiere detectar. Cuanto mayor es el TE postulado, manteniéndose lo demás constante (nivel de significación y tamaño muestral), mayor es la potencia estadística. Por otro lado, a mayor TE, la muestra necesaria para detectarlo es menor. No obstante, la mayoría de efectos buscados en investigaciones sobre personalidad, clínico-psicológicas y sociales son pequeños o moderados

(Cohen, 1988; Cooper, 1981; Cooper y Findley, 1982; Stevens, 1986). Los efectos grandes son, con frecuencia, un tema de discusión en ámbitos como la economía, la psicología experimental y la fisiológica, campos caracterizados por un buen control experimental.

A raíz de esa falta de estudios sobre TE en las ciencias del comportamiento, sería interesante que los investigadores informaran sobre los índices del TE de sus trabajos para que, así, en investigaciones futuras, haya una base de datos que permita estimar la potencia a través de, por ejemplo, procedimientos meta-analíticos (véase Cordray y Sonnefeld, 1985; Glass, McGaw y Smith, 1981; Hedges y Olkin, 1985; Holmes, 1984; Hunter, Schmidt y Jackson, 1982; Rosenthal, 1984; Rossi, 1990). Sin embargo, debe tenerse en cuenta que los estudios publicados no son representativos de todas las investigaciones llevadas a cabo (Lane y Dunlap, 1978). En efecto, aquéllas que rechazan la  $H_1$  (debido frecuentemente a una baja potencia), la mayoría de veces no son admitidas para su publicación (Greenwald, 1975; Rosenthal, 1979).

### Tipos de análisis de potencia

Dos son las formas de estimar la potencia: *a priori* y *a posteriori*. La primera alerta al investigador del tamaño de muestra necesario para una potencia adecuada. Con este fin se construyen las tablas de potencia. Por ejemplo, la tabla 2 permite hallar en que medida el estadístico *C* es capaz de detectar una tendencia en la fase de línea base de una serie temporal (Bono, 1994). En ella se muestra, para  $\alpha=0.05$ , el límite -línea discontinua- a partir del cual la potencia del estadístico *C* es superior o igual a 0.80. Lo que básicamente aporta esta tabla es el número de registros necesarios para una potencia dada. De este modo, con un TE medio (pendiente de -0.4) se requiere un tamaño muestral de 12 observaciones. Los valores de la tabla 2 no son alentadores para efectos pequeños. En concreto, se

tiene que con pendientes de -0.1 y -0.2 no existe ni un 60% de posibilidades de detectar tales pendientes, cualquiera que sea la muestra. Con tamaños del efecto grandes (pendientes de -0.7, -0.8, -0.9, y -1), son suficientes ocho registros para proporcionar una potencia igual o mayor de 0.80.

La segunda forma de estimar la potencia, *a posteriori*, es importante en la interpretación de los resultados de estudios ya acabados,

especialmente cuando son negativos (Rossi, 1990; Stevens, 1986). La mayoría de investigadores que no están familiarizados con la potencia suelen interpretar los resultados no significativos en términos de que no existe efecto alguno. Sin embargo, puede ser que el experimento posea una potencia pobre (como consecuencia de un tamaño de muestra y/o efecto pequeño) incapaz de detectar la significación estadística (Cronbach y Snow, 1977).

**Tabla 2.** Potencia del estadístico  $\bar{C}$ : series de línea base con tendencia y  $\alpha=0.05$ . (Tomado de Bono, 1994).

$n_1$	Pendiente									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8	0.050	0.130	0.200	0.395	0.580	0.720	0.800	0.875	0.930	0.925
9	0.055	0.175	0.290	0.495	0.690	0.830	0.895	0.945	0.970	1.000
10	0.085	0.165	0.430	0.615	0.770	0.915	0.960	0.990	0.985	1.000
11	0.100	0.245	0.500	0.730	0.880	0.950	1.000	1.000	0.995	1.000
12	0.115	0.295	0.640	0.885	0.935	0.985	0.990	1.000	1.000	1.000
13	0.140	0.335	0.650	0.935	0.975	0.980	1.000	1.000	1.000	1.000
14	0.150	0.405	0.755	0.920	0.990	0.995	1.000	1.000	1.000	1.000
15	0.155	0.560	0.835	0.975	1.000	1.000	1.000	1.000	1.000	1.000

En los apartados anteriores se han descrito cuatro elementos de la inferencia estadística: potencia, nivel de significación ( $\alpha$ ), tamaño de muestra ( $n$ ) y tamaño del efecto (TE). Todos ellos están mutuamente interrelacionados, de manera que cada uno es función de los otros tres. Esta relación posibilita cuatro tipos de análisis de potencia (Cohen, 1965; 1988): a) potencia en función de  $\alpha$ , TE y  $n$ ; b)  $n$  en función de  $\alpha$ , TE y potencia; c) TE en función de  $\alpha$ ,  $n$  y potencia, y d)  $\alpha$  en función de  $n$ , TE y potencia.

En el primer caso se determina la potencia, dado  $\alpha$ , TE y  $n$ . De este modo, el investigador puede decidir cambiar las especificaciones iniciales con el fin de aumentar la potencia. En el segundo tipo de análisis, el experimentador

estima un TE, establece un nivel de significación  $\alpha$  y la potencia deseada, con lo que se determina la  $n$  necesaria. Esta segunda clase de análisis es la base para la decisión del tamaño muestral a utilizar en una investigación concreta. La tercera modalidad de estudios de potencia permite encontrar el TE, dados los restantes elementos. Esta forma de análisis es de menor utilidad que las anteriores; no obstante, puede ser muy favorable en determinadas circunstancias (véase Cohen, 1970). El último modo de análisis de potencia responde a la pregunta: ¿qué  $\alpha$  es necesario para detectar un TE, con una probabilidad específica (potencia) y una  $n$  fijada?. Esta clase de análisis es muy poco común debido a la existencia de una convención de  $\alpha=0.05$ .

En el presente apartado se han descrito cuatro tipos distintos de análisis de potencia estadística. Quizá el procedimiento más frecuente sea determinar los tamaños muestrales de las investigaciones que se desean llevar a cabo. Las tablas de Cohen (1988) permiten obtener, para algunas pruebas, los dos primeros tipos de análisis y también, aunque no con tanta facilidad, los otros dos, a través de una interpolación inversa.

## Conclusiones

El estudio de la potencia de una prueba estadística es útil para la planificación y valoración de investigaciones. En efecto, las tablas de potencia permiten determinar, fundamentalmente, los tamaños muestrales de estudios que se desean llevar a cabo, así como su potencia, o bien interpretar resultados no significativos de

trabajos ya finalizados. No obstante, si el investigador no tiene una idea del grado en que la  $H_0$  puede ser falsa, la determinación de la potencia y del tamaño muestral necesario no es posible. Por lo general, se formulan hipótesis teóricas sobre el tamaño del efecto.

La potencia merece una atención especial en la decisión sobre si un experimento debe ser llevado a cabo o no. Si ya se ha decidido realizar el experimento, la potencia es importante pero no fundamental. Por otra parte, si al que hace el experimento le interesa descubrir una desviación de una cierta magnitud respecto de la  $H_0$ , entonces es imprescindible conocer la potencia con que se está trabajando. Si, por ejemplo, sólo hay una probabilidad del 15 por ciento de obtener tal diferencia, no interesa hacer el experimento. En este caso, el investigador quizá desee usar sus recursos en una investigación donde tenga más seguridad de llegar a una conclusión correcta.

## Referencias bibliográficas

- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 432-437.
- Bavry, J.L. (1991). *STAT-POWER: Statistical Design Analysis System* (2ª ed.). Chicago, IL: Scientific Software, Inc.
- Bono, R. (1994). Diseños de series temporales interrumpidas: técnicas alternativas de análisis. (Doctoral dissertation, University of Barcelona, 1994). *Dissertation Abstracts International*, 2521.
- Borenstein, M. y Cohen, J. (1988). *Statistical power analysis: A computer program*. Hillsdale, NJ: Erlbaum.
- Borenstein, M., Cohen, J., Rothstein, H.R., Pollack, S. y Kane, J.M. (1990). Statistical power analysis for one-way analysis of variance: A computer program. *Behavior Research Methods, Instruments, and Computers*, 22, 271-282.
- Brown, G.W. (1983). Errors, Type I and II. *American Journal of Disorders in Childhood*, 137, 586-591.
- Cascio, W.F. y Zedeck, S. (1983). Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Personnel Psychology*, 36, 517-526.
- Chase, L.J. y Tucker, R.K. (1976). Statistical power: derivation, development, and data-analytic implications. *The Psychological Record*, 26, 473-486.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1965). Some statistical issues in psychological research. En B.B. Wolman (Ed.), *Handbook of clinical psychology*. New York: McGraw-Hill.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.
- Cohen, J. (1970). Approximate power and sample size determination for common one-sample and two-sample hypothesis tests. *Educational and Psychological Measurement*, 30, 811-831.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Cooper, H.M. (1981). On the significance of effects and the effects of significance. *Journal of Personality and Social Psychology*, 41, 1013-1018.
- Cooper, H.M. y Findley, M. (1982). Expected effect sizes: estimates for statistical power analysis in social psychology. *Personality and Social Psychology Bulletin*, 8, 168-173.
- Cordray, D.S. y Sonnefeld, L.J. (1985). Quantitative synthesis: An actuarial base for planning impact evaluations. *New Directions for Program Evaluation*, 27, 29-48.

- Cox, D.R. (1948). Some problems connected with statistical inference. *Annals of Mathematical Statistics*, 29, 357-372.
- Cronbach, L. y Snow, R. (1977). *Aptitudes and instructional methods*. New York: Irvington Press.
- Fisher, R.A. (1935). *The design of experiments*. London: Oliver and Boyd.
- Fisher, R.A. (1950). *Contributions to mathematical statistics*. London: Chapman and Hall.
- Fisher, R.A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society, Series B*, 17, 245-251.
- Frias, D., García, J.F. y Pascual, J. (1993, Julio). Estudio de la potencia de los trabajos publicados en "Psicológica". Estimación del número de sujetos fijando  $\alpha$  y  $\beta$ . *Actas del III Simposium de Metodología de las Ciencias Sociales y del Comportamiento*, Santiago de Compostela, La Coruña.
- Gigerenzer, G. y Murray, D.J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Glass, G.V., McGaw, B. y Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Goldstein, R. (1989). Power and sample size via MS/PC-DOS computers. *American Statistician*, 43, 253-260.
- Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1-20.
- Hedges, L.V. y Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Holmes, C.T. (1984). Effect size estimation in meta-analysis. *The Journal of Experimental Education*, 52, 106-109.
- Hunter, J.E., Schmidt, F.C. y Jackson, G.B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.
- Kazdin, A.E. y Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, 57, 138-147.
- Kraemer, H.C. y Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.
- Labovitz, S. (1968). Criteria for selecting a significance level: A note on the sacredness of .05. *The American Sociologist*, 3, 220-222.
- Lane, D.M. y Dunlap, W.P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology*, 31, 107-112.
- Lipsey, M.W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Lykken, D. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151-159.
- McNemar, Q. (1960). At random: sense and nonsense. *American Psychologist*, 15, 295-300.
- Mosteller, F. y Bush, R.R. (1954). Selected quantitative techniques. En G. Lindzey (Ed.), *Handbook of social psychology: Theory and method*, Vol. 2. Reading, MA: Addison-Wesley.
- Nagel, S.S. y Neef, M. (1977). Determining an optimum level of statistical significance. En M. Guttentag y S. Saar (Eds.), *Evaluation studies review annual*, Vol 2. Beverly Hills, CA: Sage.
- Neyman, J. y Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A, 175-240.
- Neyman, J. y Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypothesis. *Transactions of the Royal Society of London Series A*, 231, 289-337.
- Neyman, J. y Pearson, E.S. (1936). Contributions to the theory of testing statistical hypotheses. *Statistical Research Memoirs*, 1, 1-37.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Pascual, J. Frias, D. y García, J. (1993, Julio). Análisis comparativo del tamaño del efecto y la potencia en función de la naturaleza del trabajo en la revista "Anuario de Psicología". *Actas del III Simposium de Metodología de las Ciencias Sociales y del Comportamiento*, Santiago de Compostela, La Coruña.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Rossi, J.S. (1990). Statistical power of psychological research: What have we gained in 20 years?. *Journal of Consulting and Clinical Psychology*, 58, 646-656.
- Rothstein, H., Borenstein, M., Cohen, J. y Pollack, S. (1990). Statistical power analysis for multiple regression/correlation: A computer program. *Educational and Psychological Measurement*, 50, 819-830.
- Runyon, R.P. (1969). Minimum sample size required to achieve power 1-B in testing for the significance of a difference in independent proportions. *Perceptual and Motor Skills*, 28, 247-250.
- Sánchez, J., Valera, A., Veladrino, A.P. y Marín, F. (1992). Un estudio de la potencia estadística en Anales de Psicología (1984-1991). *Anales de Psicología*, 8, 19-32.
- Schneider, A.L. y Darcy, R.E. (1984). Policy implications of using significance tests in evaluation research. *Evaluation Review*, 8, 573-582.
- Sedlmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies?. *Psychological Bulletin*, 105, 309-316.
- Skipper, J.K., Guenther, A.S. y Nass, G. (1967). The sacredness of .05: A note concerning the uses of statistical levels of significance in social science. *The American Sociologist*, 1, 16-18.
- Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa. *Journal of the American Statistical Association*, 54, 30-34.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. New Jersey: Erlbaum.
- Strube, M.J. (1985). Power analysis for combining significance levels. *Psychological Bulletin*, 98, 595-599.
- Tukey, J.W. (1960). Conclusions vs. decisions. *Technometrics*, 2, 423-433.

- Tullock, G. (1959). Publication decisions and tests on significance -a comment. *Journal of the American Statistical Association*, 54, 593.
- Valera, A., Sánchez, J., Velandrino, A.P. y Marín, F. (1993, Julio). Un estudio de la potencia estadística de la revista de Psicología General y Aplicada (1990-1992). *Actas del III Simposium de Metodología de las Ciencias Sociales y del Comportamiento*, Santiago de Compostela, La Coruña.

