

DTM: Determinación del tamaño de la muestra en el entorno SPSS

Dolores Frías* , Juan Pascual y José Fernando García

Universidad de Valencia

En la investigación aplicada resulta cada vez más importante determinar la potencia de las pruebas estadísticas y estimar el tamaño del efecto con el fin de determinar el tamaño de la muestra. En este trabajo se presenta un procedimiento de cálculo iterativo para su ejecución en el programa SPSS. El máximo error cometido por este procedimiento en la estimación de la potencia fue 0.001.

Palabras clave: Potencia, tamaño del efecto, tamaño de la muestra, SPSS

POTENCIA, ERROR DE TIPO I, TAMAÑO DEL EFECTO Y NÚMERO DE OBSERVACIONES

La inferencia estadística está sometida a dos posibles errores: el error de *Tipo I*, que se comete cuando el tratamiento no es efectivo, pero el investigador concluye que sí lo es; y el error de *Tipo II*, que se produce cuando el tratamiento es efectivo pero el investigador concluye que no lo es (Pascual, García y Frías, 1995). La probabilidad de cometer un error de *Tipo II* (β) es la probabilidad complementaria de $(1 - \beta)$; ésta es la potencia de la prueba estadística y está directamente relacionada con tres parámetros (Cohen, 1977; Maxwell y Delaney, 1990):

1. *La probabilidad de error de tipo I* (α): cuanto mayor es el margen de error de *Tipo I*, mayor es la potencia. Considerando los mismos datos, con un $\alpha = 0.05$ se obtiene más potencia que con un $\alpha = 0.01$.

2. *El tamaño del efecto*: cuando más grande sea el tamaño del efecto en la población de estudio, más probable es que el efecto observado en el estudio alcance el criterio de significación estadística, y mayor será la potencia.

* Correspondencia a Dolores Frías. Area de Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universitat de València. Avda. Blasco Ibáñez, 21. 46010-Valencia. E-mail: M.Dolores.Frias@uv.es

3. *El tamaño de la muestra:* conforme se incluyen más observaciones en un experimento se incrementa la probabilidad de observar el efecto experimental y, consecuentemente, se aumenta la potencia.

En el planteamiento de la investigación resulta muy sencillo comprender la relación inversa entre los dos márgenes de error. Cuanto mayor es el riesgo que asume el investigador de rechazar falsamente la hipótesis nula menor será la probabilidad de rechazarla correctamente. El investigador fija antes de iniciar la prueba de la hipótesis el riesgo de error de *Tipo I*, a no ser que especifique lo contrario el margen que convencionalmente se asume es el de 0.05 -en el caso de que la hipótesis nula sea cierta se rechazará (equivocadamente) el 5% de las veces- ¿Qué ocurre si la hipótesis nula es falsa? En este caso no resulta tan sencillo determinar cuál será la probabilidad de rechazar la hipótesis nula cuando sea falsa. Para calcular la potencia hay que tener en cuenta los tres parámetros anteriormente señalados. Aunque resulta difícil encontrar trabajos publicados que estimen la potencia.

También resulta sencillo intuir que cuanto mayor sea el tamaño del efecto que se estudie mayor será la probabilidad de detectar diferencias significativas en las muestras extraídas, por tanto, existe una relación directa y positiva entre el tamaño del efecto y la potencia de la prueba.

Pero hay un tercer factor que también incide en la potencia, el tamaño de la muestra, cuantas más observaciones se incluyan en la muestra más fácil es rechazar una hipótesis nula falsa. El tamaño de la muestra es una decisión que compete al investigador pero que en muy contadas ocasiones llega a justificarse, el número de observaciones del estudio se acostumbra a determinar en función de la tradición, de la accesibilidad de los participantes o de otras circunstancias ajenas al propio planteamiento metodológico.

PLANIFICACIÓN EXPERIMENTAL Y POTENCIA

A la hora de tomar decisiones el análisis aislado de cada uno de los tres parámetros dificulta una visión panorámica de la estrategia metodológica que se ha de seguir. El objetivo de cualquier investigador tiene que centrarse en conseguir los mínimos márgenes de error posibles para garantizar la generabilidad de sus conclusiones. El tamaño del efecto no puede variarse puesto que se trata del fenómeno que se quiere estudiar, aunque cuando se prueba una hipótesis no se conoce cual es el valor poblacional de este parámetro sino que se realiza una estimación a partir de la muestra extraída. En este sentido cabe señalar que hay que perfilar las estrategias metodológicas para conseguir mejorar la estimación mediante la elección de un diseño adecuado, la correcta operativización de las variables, etc. El tamaño del efecto no es, por tanto, un parámetro que se pueda controlar directamente aunque sí cabe tomar las precauciones necesarias para su correcta estimación. Una vez se ha estimado cual es el tamaño del efecto que se pretende detectar, el paso siguiente consiste en determinar las condiciones óptimas del experimento para garantizar que los márgenes de error de la decisión experimental sean mínimos (García, Pascual y Frías, 1995).

La primera decisión que toma el investigador consiste en fijar el margen de error de *Tipo I* que asume. Si se toman valores conservadores de la hipótesis nula (0.001 en vez de 0.05) se garantiza que es más difícil rechazar una hipótesis nula cierta, pero por contra se aumenta el riesgo de mantener hipótesis nulas falsas. En este sentido Cohen (1977, 1992) propone como un criterio válido para aumentar la potencia disminuir la probabilidad de error de *Tipo I*.

Pero la mejor propuesta para garantizar los resultados de la investigación consiste en asumir los menores márgenes posibles de error de *Tipo I* y a la vez de *Tipo II*, entonces la única estrategia posible se basa en emplear el suficiente número de observaciones. Aunque este procedimiento resulte el mejor presenta el inconveniente de que en ocasiones resulta difícil recoger una muestra amplia por cuestiones de economía. En este caso tampoco se puede recomendar que los investigadores acopien un número exagerado de observaciones con el fin de reducir los márgenes de error en su decisión.

Parece, por tanto, que la recomendación más adecuada consiste en efectuar una estimación de la potencia que garantice que en el caso de que el efecto se encuentre presente en la población se pueda rechazar la hipótesis nula con suficientes garantías. Este requisito es especialmente exigible a la investigación aplicada, puesto que en este caso no solo es necesario rechazar la hipótesis nula, sino que también se requiere que el tamaño del efecto sea lo suficientemente considerable para producir efectos que justifiquen la aplicación de los tratamientos, e incluso que permitan elegir uno de los procedimientos entre un abanico de posibilidades.

Puede ocurrir que el tamaño del efecto de un tratamiento sea de 0.001 —el 0.1% de la varianza de la variable dependiente sea atribuible a la acción del tratamiento— en este supuesto, si se hubiese empleado una muestra con 7000 observaciones se podría rechazar la hipótesis nula ¿Pero, merece la pena aplicar un tratamiento que produzca un cambio tan insignificante? Posiblemente, no. El emplear un número de observaciones muy alto puede conducir a rechazar la hipótesis nula cuando el tamaño del efecto sea muy pequeño, pero el mero hecho de rechazar la hipótesis nula no avala necesariamente la eficacia práctica del tratamiento.

Resulta más difícil estructurar una estrategia cuando se trata de valorar la magnitud del efecto en el contexto de la investigación teórica, puesto que el tamaño del efecto carece de importancia substantiva, lo único que pretende el investigador es rechazar una hipótesis nula que sea falsa, en este caso cuando el investigador no consiga rechazar la hipótesis nula siempre le queda la esperanza de aumentar el número de observaciones para intentar detectar el tamaño del efecto (Frías, Pascual y García, 1994).

CÁLCULO DE LA POTENCIA

Resulta por tanto imprescindible planificar cuál será la potencia de la prueba de la hipótesis, aunque el cálculo de este valor viene dificultado por la

complejidad de su estimación. Para facilitar este cálculo se ha desarrollado un algoritmo de cálculo ejecutable en la aplicación *SPSS para Windows*. El algoritmo estima la potencia de la *distribución F*, incrementando el número de observaciones hasta un número definido, y con un incremento controlado por el usuario. La estimación obtenida con esta función —comparándolo con los datos exactos aportados por Tiku (1967)— es prácticamente exacta, el error es inferior a una milésima.

ALGORITMO DE CÁLCULO

La potencia está calculada por la función:

$$NCDF.F(F_{crítica}, v_1, v_2, \lambda) \quad [1]$$

donde:

$$v_1 = a - 1$$

$$v_2 = N - a$$

$$\lambda = \frac{\eta^2 (N - a)}{1 - \eta^2} = F(a - 1)$$

N es el número total de observaciones

a es el número de condiciones

$F_{crítica}$ es el valor tabular de F para v_1 y v_2 grados de libertad

$NCDF.F$ calcula la función de distribución no-centrada del estadístico F . Esta función proporciona la probabilidad de que una variable aleatoria con la distribución especificada sea menor que $F_{crítica}$, el primer argumento. Los argumentos subsiguientes son los parámetros de la distribución. Por tanto, $NCDF.F$ proporciona la probabilidad acumulada de que una distribución F no-centrada, con los grados de libertad v_1 y v_2 , y descentramiento λ , sea menor que $F_{crítica}$. La función facilita, por tanto, el error de *Tipo II*, siendo el valor de la potencia $(1 - \beta)$ el complementario de esta cantidad.

El valor de $F_{crítica}$ también se obtiene a partir de una función definida por este mismo programa.

$$IDF.F(1 - \alpha, v_1, v_2) \quad [2]$$

$IDF.F$ devuelve el valor de la distribución F con v_1 y v_2 grados de libertad, para una probabilidad acumulada igual a p . Para obtener el valor, a partir del margen de error de *Tipo I* se introduce $1 - \alpha$, el complementario de la parte derecha de la distribución.

Precisión. Esta función proporciona un valor de la potencia muy aproximado al valor exacto calculado por Tiku (1967). Este autor utilizaba como parámetro de no centralidad:

$$\phi = f_d \sqrt{n} \quad [3]$$

Otros autores han diseñado procedimientos de cálculo para aproximar el cálculo de la potencia sin necesidad de tener que consultar estas extensas tablas. Severo y Zelen (1960) aproximaron el valor de la potencia de la prueba F a partir de la distribución normal, la fórmula proporcionada por estos autores cuenta con la ventaja de poder resolverse con una calculadora y las tablas de la distribución normal reducida (Pascual, García y Frías, 1995). Con la proliferación de los ordenadores han surgido multitud de procedimientos para el cálculo que cuentan con la ventaja de estimar la potencia fijando el resto de parámetros (Goldstein, 1989; Graf y Alf, 1990). Borenstein, Cohen, Rothstein, Pollack y Kane (1990) desarrollan otro programa para calcular la potencia a partir del índice f . En general, el parámetro de no centralidad más utilizado es:

$$\lambda = f_d^2 N \quad [4]$$

la equivalencia entre ambos es muy sencilla de establecer en un diseño equilibrado:

$$\lambda = f_d^2 N = \frac{\phi^2}{n} N = \phi^2 \frac{N}{n} = \phi^2 a = \phi^2 (v_1 + 1)$$

A partir de esta transformación se ha calculado la potencia estimada por los tres procedimientos reseñados y se ha estimado el error que cada uno de ellos cometía en relación a los valores exactos de Tiku (ver Tabla 1).

El mejor procedimiento se obtiene estimando la potencia a partir de la función $NCDF.F$ del programa SPSS.

LISTADO DE INSTRUCCIONES

Este programa está diseñado para transcribirlo en una hoja de *Sintaxis* del SPSS y lanzarlo cuando sea necesario, sin necesidad de reescribirlo cada vez (Figura 1). No hay que introducir los números de línea, puesto que la sintaxis del programa no los acepta. También hay que terminar cada una de las sentencias con un punto, para indicar el final de cada instrucción, tal y como figura en el listado. Los espacios en blanco y los guiones también deben escribirse.

La primera línea operativa del programa {5} abre un fichero. En la segunda línea operativa {6} se inicia la generación del fichero de datos mediante el comando *INPUT PROGRAM..*

Tabla 1. Estimación de la potencia según procedimiento de Tiku, Borenstein, Severo-Zelen y SPSS.

α	Tiku					Borenstein		Severo-Zelen		SPSS	
	v_1	v_2	ϕ	λ'	$1-\beta$	$1-\beta$	D1	$1-\beta$	D2	$1-\beta$	D3
0.01	1	20	0.50	1	0.028	0.021	0.007	0.026	0.002	.028	.000
0.01	1	20	1.00	2	0.101	0.096	0.005	0.095	0.006	.101	.000
0.01	1	20	2.00	8	0.508	0.515	-0.007	0.498	0.010	.507	.001
0.01	1	20	3.00	18	0.904	0.909	-0.005	0.915	-0.011	.904	.000
0.01	3	20	0.50	1	0.027	0.024	0.003	0.027	0.000	.027	.000
0.01	3	20	1.00	4	0.113	0.112	0.001	0.110	0.003	.113	.000
0.01	3	20	2.00	16	0.653	0.653	0.000	0.650	0.003	.653	.000
0.01	3	20	3.00	36	0.979	0.980	-0.001	0.982	-0.003	.979	.000
0.01	9	20	0.50	3	0.029	0.030	-0.001	0.029	-0.000	.029	.000
0.01	9	20	1.00	10	0.159	0.159	0.000	0.157	0.002	.159	.000
0.01	9	20	2.00	40	0.864	0.863	0.001	0.866	-0.002	.864	.000
0.01	9	20	3.00	90	1.000	0.999	0.001	1.000	0.000	.999	.001
0.05	1	20	0.50	1	0.103	0.101	0.002	0.098	0.005	.103	.000
0.05	1	20	1.00	2	0.270	0.280	-0.010	0.257	0.013	.270	.000
0.05	1	20	2.00	8	0.768	0.783	-0.016	0.778	-0.010	.767	.001
0.05	1	20	3.00	18	0.951	0.983	-0.002	0.988	-0.007	.981	.000
0.05	3	20	0.50	1	0.104	0.104	0.000	0.102	0.002	.104	.000
0.05	3	20	1.00	4	0.300	0.306	-0.006	0.293	0.007	.300	.000
0.05	3	20	2.00	16	0.674	0.877	-0.003	0.881	-0.007	.874	.000
0.05	3	20	3.00	36	0.995	0.998	0.000	0.999	-0.001	.998	.000
0.05	9	20	0.50	3	0.114	0.116	-0.002	0.114	-0.000	.114	.000
0.05	9	20	1.00	10	0.391	0.393	-0.002	0.388	0.003	.391	.000
0.05	9	20	2.00	40	0.974	0.975	-0.001	0.976	-0.002	.974	.000
0.05	9	20	3.00	90	1.000	1.000	0.000	1.000	0.000	1.000	.000
$\Sigma (1-\beta) / 12$						$-0.036/12 = -0.003$		$0.013/12 = 0.001$		$0.003/12 < 0.001$	
$\Sigma 1-\beta / 12$						$0.076/12 = 0.006$		$0.099/12 = 0.008$		$0.003/12 < 0.001$	

La línea {7} permite controlar el número de ensayos que se van a realizar (está definido en 30) y el incremento que se quiere realizar en cada bucle (1). Estos valores pueden cambiarse en función del tamaño del efecto observado. En el caso de que el tamaño del efecto sea muy pequeño se pueden incrementar cada vez 10 o 20 observaciones para que de esta forma se aumente substancialmente la potencia. Cuando el tamaño del efecto es reducido, pudieran en ocasiones necesitarse 1000 observaciones o más, y si hay que ir incrementando de uno en uno el resultado pudiera ser muy extenso. En este programa resulta muy sencillo variar estos parámetros.

En las líneas {11-14} se definen los valores necesarios para iniciar la estimación: El número de niveles de la variable independiente {a ... ≡ a}, el número de observaciones de las que partimos {N ... ≡ #N}, el margen de error de Tipo I que asumimos { α ... ≡ Alfa} y el tamaño del efecto que presumimos en la población { η^2 ... ≡ Eta2}. Los valores hay que modificarlos en función de cuales sean los parámetros de nuestra investigación.

Las líneas {18-19} van incrementando el tamaño de la muestra inicial {#N} en función del valor inicial de la variable contadora {#I}. El resultado se registra en {N}. En el primer caso el valor de {N} corresponde al inicial {#N}.

```

1.  * -----
2.  *                               D T M.
3.  *       DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA.
4.  * -----
5.  NEW FILE.
6.  INPUT PROGRAM.
7.  LOOP #I = 1 TO 30 BY 1.
8.  - * -----
9.  - *
10. - * -----
11. - COMPUTE a = 2.           /* niveles variable independiente
12. - COMPUTE #N = 4.         /* número inicial de observaciones
13. - COMPUTE Alfa = .0500.  /* Error de Tipo I
14. - COMPUTE Eta2 = 0.30.   /* Tamaño del efecto
15. - * -----
16. - *       Aumento del tamaño de la muestra.
17. - * -----
18. - IF (#I = 1) N = #N.
19. - IF (#I > 1) N = #N + #I.
20. - * -----
21. - *       Cálculo de la Potencia.
22. - * -----
23. - COMPUTE v1 = a - 1.     /* gl del numerador
24. - COMPUTE v2 = N - a.     /* gl del denominador
25. - COMPUTE F = ((Eta2 * (N - a)) / ((1 - Eta2) * v1)).
26. - COMPUTE Lambda = F * v1.
27. - COMPUTE Fcrit= IDF.F(1 - Alfa, v1, v2).
28. - COMPUTE Potencia = 1 - NCDF.F(Fcrit, v1, v2, Lambda).
29. - END CASE.
30. END LOOP.
31. END FILE.
32. END INPUT PROGRAM.
33. EXECUTE.
34. FORMATS Alfa Eta2 F Fcrit Potencia (F8.3) a v1 v2 N (F8.0).
35. LIST
    VARIABLES= a Alfa Eta2 N v2 F Lambda Fcrit Potencia
    /FORMAT= WRAP UNNUMBERED.

```

Figura 1. Listado del programa DTM.

Las líneas {23-28} calculan los grados de libertad, el valor de F estimado a partir del tamaño del efecto y los grados de libertad derivados del número de condiciones experimentales y observaciones, el valor crítico de F para los grados de libertad y el error de Tipo I determinado, y finalmente, se calcula la potencia a partir de los parámetros anteriores.

La línea {29} determina la finalización de cada paso, la {30} cierra el bucle, la {29} determina el final del proceso de cálculo y ordena al programa que cree el fichero de resultados. La ejecución de esta operación se dictamina mediante la sentencia *EXECUTE*.

Finalmente, la línea {35} lista las variables y los correspondientes valores de cada una de las simulaciones. En la hoja de datos del programa también se han salvado los resultados de las operaciones.

DISCUSIÓN

Sabemos por Sedlmeier y Gigerenzer (1989) que los estudios acerca de la potencia han tenido poca repercusión sobre la conducta de los investigadores, que no han modificado sus hábitos de investigación. Aparte de las dificultades conceptuales y la poca tradición, la escasa relevancia que han tenido los programas informáticos que se han desarrollado puede estar en que se programaron en lenguajes extraños al entorno operativo del investigador. El programa DTM ha sido diseñado para su ejecución con el propio programa SPSS, la flexibilidad de DTM es muy alta puesto que permite variar todos los parámetros de cálculo directamente consiguiendo una precisión adecuada (el máximo error detectado es de 0.001). Esperamos facilitar con nuestra aportación el proceso de análisis de la potencia de manera que toda la investigación experimental sea diseñada de acuerdo con el máximo rigor formal.

ABSTRACT

DTM: Determining the sample size in SPSS environment. In the applied research it is increasingly important to calculate the power of the statistic tests and to estimate the effect's size in order to determine the sample's size. In this work we present an iterative calculation procedure that can be used with the SPSS program. The maximum error obtained by this procedure during power estimation was 0.001.

Key words: power, effect size, sample size, SPSS.

BIBLIOGRAFÍA

- Borenstein, M., Cohen, J., Rothstein, H. R., Pollack, S., & 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.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (ed. rev.)*. New York: Academic Press.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, **112**, 155-159.
- Frías, M. D., Pascual, J., & García, J. F. (1994b, July). *Planificación metodológica de la recherche en psychologie appliquée*. Comunicación presentada en el 23rd International Congress of Applied Psychology, Madrid, Spain.
- García, J. F., Pascual, J., & Frías, M. D. (1994). Determinación del tamaño de la muestra en diseños univariados. En C. Arce y G. Seoane, *III Simposium de Metodología de las Ciencias Sociales y del Comportamiento* (pp. 1073-1082). Santiago de Compostela: Servicio de Publicaciones e Intercambio Científico.
- Goldstein, R. (1989). Power and sample size via MS/PC-DOS computers. *American Statistician*, **43**, 253-260.

- Graf, R. G., & Alf, E. F. (1990). Basic programs to determine sample size and exact power in ANOVA designs. *Educational and Psychological Measurement*, **50**, 117-121.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analysis data. A model comparison perspective*. Belmont, CA: Wadsworth Publishing Company.
- Pascual, J., García, J. F., & Frías, M. D. (1995). *El diseño y la investigación experimental en psicología*. Valencia: CSV.
- Severo, N. C., & Zelen, M. (1960). Normal approximation to the chi-square and non-central *F* probability functions. *Biometrika*, **47**, 411-416.
- Sedlmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, **105**, 309-316.
- Tang, P. C. (1938). The power function of the analysis of variance test with table and illustrations of their use. *Statistical Research Memoirs*, **2**, 126-149.
- Tiku, M. L. (1967). Tables of the power of the F-test. *Journal of the American Statistical Association*, **62**, 525-539.

(Revisión aceptada: 6/3/96)