

Cálculo del poder estadístico de un estudio

Pértegas Díaz, S. spertega@canalejo.org, Pita Fernández, S. spita@canalejo.org
Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo.
A Coruña (España). Actualización 20/01/2003.

Las pruebas de contraste de hipótesis y el concepto de poder estadístico

El análisis de estudios clínico-epidemiológicos con frecuencia exige la comparación entre varios tratamientos o entre diferentes grupos de sujetos con respecto a una respuesta de interés. Por ejemplo, puede interesarnos comparar la eficacia de un nuevo fármaco frente a la de otro estándar en el tratamiento de una determinada patología, o bien establecer la asociación entre la exposición a un factor de riesgo y el desarrollo de cierta enfermedad. Problemas de este tipo pueden plantearse como un *contraste de hipótesis*, de forma que la hipótesis que se contrasta es la de que no existen diferencias entre ambos grupos o tratamientos (*hipótesis nula*), frente a la *hipótesis alternativa* de que estos sean realmente diferentes.

Una *prueba de contraste de hipótesis* o de *significación estadística* calcula la probabilidad de que los resultados obtenidos en una investigación puedan ser debidos al azar en el supuesto de que la hipótesis nula sea cierta, es decir, bajo el supuesto de que no existan diferencias entre ambos grupos. Esta probabilidad es el *grado de significación estadística* o valor de p . Basándose en esta probabilidad, se decidirá rechazar o no la hipótesis nula. Así, cuanto menor sea el valor de p , menor será la probabilidad de que los resultados obtenidos se deban al azar y mayor evidencia habrá en contra de la hipótesis nula. Si dicha probabilidad es menor que un valor de p fijado previamente (habitualmente se toma $p < 0,05$), la hipótesis nula se rechazará. Así, cuando el valor de p está por debajo de 0,05, se dirá que el resultado es *estadísticamente significativo* y será *no significativo* en cualquier otro caso.

Así pues, a partir de los resultados de un estudio, puede llegarse a diferentes conclusiones ([Tabla 1](#)). En primer lugar, puede concluirse que existen diferencias entre los grupos que se comparan cuando realmente las hay. Asimismo, puede concluirse que no hay diferencias cuando éstas no existen. En ambos casos, no se comete ningún error. Sin embargo, también se puede concluir que existen diferencias cuando de hecho no las hay. Es decir, puede rechazarse la hipótesis nula cuando en realidad es verdadera. Si esto ocurre, se comete un error de tipo I o error α . La probabilidad de cometer un error de este tipo es lo que mide precisamente el *grado de significación p*.

En algunas ocasiones, por el contrario, los resultados de un estudio no son significativos. Se habla entonces de “*estudios negativos*”. No obstante, la ausencia de significación estadística no implica necesariamente que no exista relación entre el factor de estudio y la respuesta. Puede ocurrir, que aún existiendo tal asociación o una diferencia clínicamente relevante, el estudio haya sido incapaz de detectarla como estadísticamente significativa. En estudios de este tipo se concluirá que no existen diferencias cuando realmente sí las hay. Este error se conoce como *error de tipo II*. La probabilidad de cometer un error de este tipo suele denotarse por β y su complementario, $1-\beta$, es lo que se conoce como *poder estadístico* o *potencia estadística*¹⁻³.

En definitiva, el *poder estadístico* representa la probabilidad de rechazar la hipótesis nula cuando es realmente falsa. Es decir, representa la capacidad de un test para detectar como estadísticamente significativas diferencias o asociaciones de una magnitud determinada.

Factores que influyen en el poder estadístico de un estudio.

El *poder estadístico* de un estudio depende de diferentes factores, como³:

- *El tamaño del efecto a detectar*, es decir, la magnitud mínima de la diferencia o asociación entre los grupos que se considera clínicamente relevante. Cuanto mayor sea el tamaño del efecto que se desea detectar, mayor será la probabilidad de obtener hallazgos significativos y, por lo tanto, mayor será el poder estadístico.

- La *variabilidad* de la respuesta estudiada. Así, cuanto mayor sea la variabilidad en la respuesta, más difícil será detectar diferencias entre los grupos que se comparan y menor será el poder estadístico de la investigación. De ahí que sea recomendable estudiar grupos lo más homogéneos posibles.
- El *tamaño de la muestra* a estudiar. Cuanto mayor sea el tamaño muestral, mayor será la potencia estadística de un estudio. Es por ello que en los estudios con muestras muy grandes se detectan como significativas diferencias poco relevantes, y en los estudios con muestras menores es más fácil obtener resultados falsamente negativos.
- El *nivel de significación estadística*. Si se disminuye el valor de α también se disminuye el poder de la prueba. Es decir, si disminuimos la probabilidad de cometer un error de tipo I aumentamos simultáneamente la probabilidad de un error de tipo II, por lo que se trata de encontrar un punto de “equilibrio” entre ambas. Habitualmente se trabaja con un nivel de significación del 95% ($\alpha = 0.05$), por lo que el equilibrio hay que encontrarlo finalmente entre el tamaño de la muestra que es posible estudiar y el poder que se quiere para el estudio.

Los cuatro factores anteriores, junto con el poder estadístico, forman un sistema cerrado. De este modo, una vez fijados tres de ellos, el cuarto queda completamente determinado.

Cálculo del poder estadístico de un estudio.

A la hora de diseñar una investigación, es importante determinar si dicho estudio alcanzará una precisión suficiente. En anteriores trabajos se ha mostrado cómo calcular el tamaño muestral necesario para alcanzar un determinado poder estadístico en diferentes tipos de diseño⁴⁻⁷. Generalmente, se suele trabajar con un poder en torno al 80% o al 90%. Con frecuencia, sin embargo, las condiciones en las que se lleva a cabo una investigación son diferentes de las que se habían previsto en un principio. En consecuencia, y a la vista de hallazgos no significativos, es recomendable evaluar de nuevo a posteriori su potencia con el fin de discernir si el estudio carece del poder necesario para detectar una diferencia relevante o bien si realmente puede no existir tal diferencia.

En la [Tabla 2](#) se muestran las fórmulas necesarias para el cálculo del poder estadístico en función de la naturaleza de la investigación. Estas fórmulas permiten obtener un valor $z_{1-\beta}$ a partir del cual se puede determinar el poder asociado recurriendo a las tablas de la distribución normal. En la [Tabla 3](#) se muestra la correspondencia entre algunos valores de $z_{1-\beta}$ y el poder estadístico asociado. Sin embargo, y aunque dichas fórmulas nos permitirían analizar el poder estadístico en diferentes tipos de diseño, puede resultar más sencillo disponer de algún software específico con el que poder realizar dichos cálculos⁸.

Ejemplo 1.

Ilustremos el proceso del cálculo de la potencia mediante un ejemplo. Supongamos que se quiere llevar a cabo un ensayo clínico para comparar la efectividad de un nuevo fármaco con la de otro estándar en el tratamiento de una determinada enfermedad. Al inicio del estudio, se sabe que la eficacia del tratamiento habitual está en torno al 40%, y se espera que con el nuevo fármaco la eficacia aumente al menos en un 15%. El estudio se diseñó para que tuviese un poder del 80%, asumiendo una seguridad del 95%. Esto implica que son necesarios 173 pacientes en cada uno de los grupos para llevar a cabo la investigación. Tras finalizar el estudio, sólo fue posible tratar con cada uno de los fármacos a 130 pacientes en cada grupo en lugar de los 173 pacientes estimados inicialmente. Al realizar el análisis estadístico, se objetivó que no hay diferencias significativas en la efectividad de ambos tratamientos. A partir de las fórmulas de la [Tabla 2](#), podemos calcular cuál ha sido finalmente el poder del estudio. Aplicando la fórmula para el cálculo del poder estadístico de comparación de dos proporciones ante un planteamiento unilateral se obtiene:

$$\left. \begin{array}{l} p_1 = 0.4 \\ p_2 = 0.55 \end{array} \right\} \Rightarrow p = 0.475 \\
 \left. \begin{array}{l} n = 130 \\ \alpha = 0.05 \Rightarrow z_{1-\alpha} = 1.645 \end{array} \right\} \Rightarrow z_{1-\beta} = \frac{|p_1 - p_2| \sqrt{n} - z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}} = \\
 = \frac{|0.4 - 0.55| \sqrt{130} - 1.645 \cdot \sqrt{2 \cdot 0.475 \cdot (1 - 0.475)}}{\sqrt{0.4(1-0.4) + 0.55(1-0.55)}} = 0.467$$

A partir de la [Tabla 3](#), podemos determinar que un valor de $z_{1-\beta} = 0.467$ corresponde a un poder en torno al 65%-70%. Utilizando las tablas de la distribución normal, se sabe que el poder es del 68%, es decir, el estudio tendría un 68% de posibilidades de detectar una mejora en la eficacia del tratamiento del 15%.

Utilizando la fórmula anterior, podría obtenerse un gráfico como en el que se muestra en la [Figura 1](#), en la que, para este ejemplo, se estima el poder estadístico del estudio en función del tamaño de la muestra estudiada y la magnitud del efecto a detectar. Así, puede concluirse que de haber estudiado 130 pacientes por grupo, se obtiene una potencia de sólo el 36.6% para detectar una diferencia mínima del 10%, una potencia del 68% para detectar una diferencia del 15% y de un 90.2% para una diferencia del 20%. Este tipo de gráficos resulta muy útil tanto en la fase de diseño de un estudio como a la hora de valorar a posteriori el poder de una investigación.

Ejemplo 2.

De modo análogo, supongamos que se quiere llevar a cabo un estudio de casos y controles para estudiar la posible asociación entre la presencia de cardiopatía isquémica y el hábito de fumar. De acuerdo con estudios previos, se cree que la incidencia de cardiopatía puede ser hasta 2 veces más alta entre los fumadores, y se asume que la frecuencia de exposición entre los controles será de un 40%. Debido a ciertas limitaciones, sólo es posible para el investigador incluir en el estudio a 100 pacientes con cardiopatía isquémica (casos). Utilizando de nuevo las fórmulas de la [Tabla 2](#), con un planteamiento bilateral y una seguridad del 95%:

$$\left. \begin{array}{l} OR = 2 \\ p_2 = 0.40 \end{array} \right\} \Rightarrow p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2} = \frac{2 \times 0.40}{(1-0.40) + 2 \times 0.40} = 0.5714 \\
 p = \frac{p_1 + p_2}{2} = 0.486 \\
 n = 100 \\
 \alpha = 0.05 \Rightarrow z_{1-\alpha/2} = 1.96 \\
 c = 1
 \end{array} \right\} \Rightarrow \\
 \Rightarrow z_{1-\beta} = \frac{|p_1 - p_2| \sqrt{cn} - z_{1-\alpha} \sqrt{(c+1)p(1-p)}}{\sqrt{cp_1(1-p_1) + p_2(1-p_2)}} = \frac{|0.5714 - 0.40| \sqrt{100} - 1.96 \cdot \sqrt{2 \cdot 0.486 \cdot (1 - 0.486)}}{\sqrt{0.5714(1-0.5714) + 0.40(1-0.40)}} = 0.472$$

Recurriendo de nuevo a las tablas de la distribución normal, se obtiene para un valor $z_{1-\beta} = 0.472$ un poder del 68.17%.

Con el fin de mejorar el poder del estudio, los investigadores se plantean reclutar un mayor número de controles que de casos. En la [Figura 2](#) se muestra para el ejemplo anterior el poder de la investigación en

función del número de casos y controles estudiados. Como se puede observar, la ganancia en el poder disminuye rápidamente, y es prácticamente nula cuando la relación entre el número de controles y casos es 4:1. Esto se verifica en cualquier estudio de casos y controles³. En particular, para el ejemplo previo, si se estudiaran 100 casos y 200 controles se alcanzaría un poder del 80.28%. Si se incluyesen 100 casos y 300 controles, el poder sería de un 84.69%. Con 400 controles el poder aumentaría sólo a un 86.89% y con 500 a un 88.19%. Con lo cual claramente es ineficiente el incluir más de 4 controles por caso ya que no lograríamos un incremento relevante del poder estadístico.

El análisis adecuado del poder estadístico de una investigación, que es en definitiva la capacidad que tiene el estudio para encontrar diferencias si es que realmente las hay, es un paso fundamental tanto en la fase de diseño como en la interpretación y discusión de sus resultados. A la hora del diseño, por tanto, debe establecerse la magnitud mínima de la diferencia o asociación que se considere de relevancia clínica, así como el poder estadístico que se desea para el estudio y, de acuerdo con ello, calcular el tamaño de la muestra necesaria. Tras realizar el análisis estadístico, cuando se dice que no existe evidencia de que A se asocie con B o sea diferente de B, deberemos cuestionarnos antes de nada si la ausencia de significación estadística indica realmente que no existe una diferencia o asociación clínicamente relevante, o simplemente que no se dispone de suficiente número de pacientes para obtener hallazgos significativos. Tanto si los hallazgos son estadísticamente significativos como si no lo son, la estimación de intervalos de confianza pueden también facilitar la interpretación de los resultados en términos de magnitud y relevancia clínica, proporcionándonos una idea de la precisión con la que se ha efectuado al estimación, de la magnitud y de la dirección del efecto⁹⁻¹⁰. De este modo, los intervalos de confianza nos permiten tener una idea acerca del poder estadístico de un estudio y, por tanto, de la credibilidad de la ausencia de hallazgos significativos.

Bibliografía

1. Altman D.G. Practical Statistics for Medical Research. London: Chapman & Hall; 1991.
2. Kelsey J.L., Whittemore A.S., Evans A., Thompson W.D. Methods in Observational Epidemiology. 2nd ed. New York: Oxford University Press; 1996.
3. Argimon Pallás J.M., Jiménez Villa J. Métodos de investigación clínica y epidemiológica. 2^a ed. Madrid: Ediciones Harcourt; 2000.
4. Pita Fernández S. Determinación del tamaño muestral. Cad Aten Primaria 1996; 3: 138-141. [[Texto completo](#)]
5. Pértega Díaz S, Pita Fernández S. Cálculo del tamaño muestral para la determinación de factores pronósticos. Cad Aten Primaria 2002; 9: 30-33. [[Texto completo](#)]
6. Pértega Díaz S, Pita Fernández S. Cálculo del tamaño muestral en estudios de casos y controles. Cad Aten Primaria 2002; 9:148-150. [[Texto completo](#)]
7. Pértega Díaz S, Pita Fernández S. Determinación del tamaño muestral para calcular la significación del coeficiente de correlación lineal. Cad Aten Primaria 2002; 9: 209-211. [[Texto completo](#)]
8. Thomas L, Krebs CJ. A review of Statistical power analysis software. Bulletin of the Ecological Society of America 1997; 78 (2): 126-139.
9. Braitman LE. Confidence intervals assess both clinical significance and statistical significance. Ann Intern Med. 1991; 114 (6): 515-7. [[Medline](#)]
10. Argimon JM. El intervalo de confianza: algo más que un valor de significación estadística. Med Clin (Barc) 2002; 118(10): 382-384. [[Medline](#)]

Tabla 1. Posibles conclusiones tras una prueba estadística de contraste de hipótesis.			
		Resultado de la prueba	
		Asociación o diferencia significativa	Asociación o diferencia no significativa
Realidad	Existe asociación o diferencia	No error $(1 - \beta)$	Error de tipo II β
	No existe asociación o diferencia	Error de tipo I α	No error $(1 - \alpha)$

α = probabilidad de cometer un error de tipo I.

β = probabilidad de cometer un error de tipo II.

Tabla 2. Fórmulas para el cálculo del poder estadístico para diferentes tipos de diseño.			
	Test unilateral		Test bilateral
Comparación de dos proporciones	$z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$		$z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - z_{1-\alpha/2} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
Comparación de dos media	$z_{1-\beta} = \sqrt{\frac{n}{2}} \frac{d}{S} - z_{1-\alpha}$		$z_{1-\beta} = \sqrt{\frac{n}{2}} \frac{d}{S} - z_{1-\alpha/2}$
Estimación de un OR en estudios de casos y controles	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{nc} - z_{1-\alpha} \sqrt{(c+1)p(1-p)}}{\sqrt{c \cdot p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \cdot n$		$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{nc} - z_{1-\alpha/2} \sqrt{(c+1)p(1-p)}}{\sqrt{c \cdot p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \cdot n$
Estimación de un RR	$p_1 = RR \cdot p_2$ $z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$		$p_1 = RR \cdot p_2$ $z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - z_{1-\alpha/2} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
Estimación de un coeficiente de correlación lineal	$z_{1-\beta} = \sqrt{n-3} \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - z_{1-\alpha}$		$z_{1-\beta} = \sqrt{n-3} \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - z_{1-\alpha/2}$

- n = Tamaño muestral. En un estudio de casos y controles, n es el número de casos.
- p_1 = En un estudio transversal o de cohortes, proporción de expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de casos expuestos.
- p_2 = En un estudio transversal o de cohortes, proporción de no expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de controles expuestos.
- $p = \frac{p_1 + p_2}{2}$
- d = Valor mínimo de la diferencia a detectar entre dos medias
- S^2 = Varianza en el grupo control o de referencia

- c = Número de controles por caso
- m = En un estudio de casos y controles, número de controles
- OR = Valor aproximado del odds ratio a detectar
- RR = Valor aproximado del riesgo relativo a detectar
- r = Magnitud del coeficiente de correlación a detectar

Tabla 3. Valores de $z_{1-\alpha}$, $z_{1-\alpha/2}$ y $z_{1-\beta}$ más frecuentemente utilizados.

Seguridad	α	Test unilateral $z_{1-\alpha}$	Test bilateral $z_{1-\alpha/2}$
80 %	0,200	0,842	1,282
85 %	0,150	1,036	1,440
90 %	0,100	1,282	1,645
95 %	0,050	1,645	1,960
97,5 %	0,025	1,960	2,240
99 %	0,010	2,326	2,576
Poder estadístico	$1-\beta$	β	$z_{1-\beta}$
99 %	0,99	0,01	2,326
95 %	0,95	0,05	1,645
90 %	0,90	0,10	1,282
85 %	0,85	0,15	1,036
80 %	0,80	0,20	0,842
75 %	0,75	0,25	0,674
70 %	0,70	0,30	0,524
65 %	0,65	0,35	0,385
60 %	0,60	0,40	0,253
55 %	0,55	0,45	0,126
50 %	0,50	0,50	0,000

