

Nota Estadística

Potencia estadística en el diseño de estudios clínicos

Jorge Camacho-Sandoval

Las notas estadísticas anteriores hicieron referencia, entre otras cosas, a las pruebas de hipótesis para decidir, por ejemplo, si la diferencia entre el promedio del grupo tratado y del grupo control, era estadísticamente significativa, es decir, si se podía atribuir esa diferencia al efecto del tratamiento y no al azar.

Es importante analizar con mayor detalle los posibles resultados que se pueden obtener al realizar una prueba de hipótesis, como se ilustra en el Cuadro 1. Es posible concluir que existen diferencias significativas cuando realmente las hay, es decir se llega a una conclusión correcta. También es posible concluir que no existen diferencias significativas cuando realmente no existen, es decir, nuevamente una decisión correcta.

Cuadro 1. Posibles conclusiones en pruebas de significación estadística (Lipsey, 1990)		
Conclusión de la prueba estadística	Situación real en la población	
	Tratamiento y control SI difieren	Tratamiento y Control NO difieren
Diferencias significativas (Rechazar H_0)	Conclusión correcta Probabilidad = $1-\beta$ (Potencia)	Error tipo I Probabilidad = α
Diferencias NO significativas (Aceptar H_0)	Error tipo II Probabilidad = β	Conclusión correcta Probabilidad = $1-\alpha$

Existen otras dos posibilidades, una es concluir que existen diferencias significativas cuando realmente no las hay, o sea rechazar una hipótesis nula verdadera, lo que constituye un error Tipo I. La última posibilidad es concluir que no existen diferencias cuando éstas realmente existen, situación identificada a veces como “falsos negativos”, es decir aceptar una hipótesis nula inválida, lo que constituye un error Tipo II.

El interés del investigador es reducir al mínimo posible la probabilidad de incurrir en ambos tipos de error. La probabilidad de error tipo I es α , es decir el nivel de significación. Ésta probabilidad la determina el investigador, como límite superior, antes de proceder al análisis y a menudo su valor se fija en 0.05 o 5%. En el caso del error tipo II, solo puede determinarse hasta que se conoce el efecto del tratamiento, por lo que en el proceso de diseño del estudio se deben utilizar estimaciones. Por otra parte, ambos tipos de error están inversamente relacionados, de manera que si se reduce el error tipo I, se incrementa el error tipo II. La única manera de reducir ambos tipos de error es incrementando el tamaño de muestra.

Entonces, se puede limitar la probabilidad de error tipo I estableciendo el nivel de significancia a un nivel adecuado, como 0.05 o 0.01. ¿Cómo controlar la probabilidad de error tipo II? ¿Qué tamaño de muestra utilizar para controlar el error tipo II?

En realidad lo que se hace es utilizar la probabilidad complementaria, es decir, la probabilidad de encontrar diferencias de cierta magnitud cuando éstas realmente existen, lo que se conoce como potencia estadística ($1 - \beta$). Se debe llevar a cabo un estudio si el mismo tiene una alta potencia, o sea, una alta probabilidad de declarar significativas diferencias de cierta magnitud, si realmente existen. Realizar estudios o experimentos con baja potencia tiene implicaciones bioéticas, ya que supone “*exponer a los participantes a los riesgos y cargas de la investigación clínica*”, en estudios que carecerían de “*utilidad científica*” (Latour-Pérez, 2002).

Profesor, Maestría en
Epidemiología, Posgrado en
Ciencias Veterinarias, UNA.
Correo electrónico:
jcamacho@ice.co.cr

ISSN 0001-6002/2007/49/4/203-204
Acta Médica Costarricense, ©2007
Colegio de Médicos y Cirujanos

Algunos autores, Cohen (1992) por ejemplo, consideran alta una potencia igual o superior a 0.80, lo que se debe tener en cuenta al diseñar un estudio clínico y, particularmente, al definir el tamaño de muestra. La potencia estadística depende de varios factores, entre ellos el tamaño del efecto a detectar o la magnitud mínima de la diferencia que se considera clínicamente importante, la variabilidad de la respuesta estudiada, el nivel de significancia escogido (α) y el tamaño de muestra. En la figura 1 se muestran los tamaños de muestra necesarios para alcanzar distintos niveles de potencia en función de efectos de magnitud variable, expresados en términos de desviación estándar de la variable respuesta de interés, para un nivel de significancia, $\alpha = 0.05$, al comparar los promedios del grupo tratado y el grupo control.

Note que todas las líneas muestran una tendencia creciente en función de la potencia, es decir, cualquiera sea la magnitud del efecto, al incrementarse el nivel de potencia deseado se incrementa el tamaño de muestra. Por otra parte, para un nivel de potencia determinado, conforme se reduce la magnitud del efecto se incrementa el tamaño de muestra.

Ramírez-Figueroa y Sotres-Ramos (2005) proponen un ejemplo en el que se compara la eficacia antihipertensiva de monoxidin y captotril. Se estableció como efecto significativo una diferencia de 8 mmHg y se estimó una desviación estándar de 10 mmHg, por lo que la magnitud del efecto estandarizado fue $d = 8/10 = 0.8$. Se estableció además un nivel de significancia, α , de 0.05. ¿Qué tamaño de muestra se requiere para alcanzar una potencia de 0.9?

En la figura ese valor se obtiene proyectando al eje vertical el punto en que una línea perpendicular al valor de potencia 0.9 corta la curva correspondiente al tamaño de efecto 0.8, como lo indican las flechas. Ese valor corresponde a un tamaño de muestra de 55. Si se quisiera un tamaño de efecto menor, por ejemplo 0.6, el tamaño de muestra sería cercano a 95.

En resumen, es importante tener en cuenta la potencia estadística cuando se diseña un estudio, de manera que el tamaño de muestra utilizado garantice una elevada

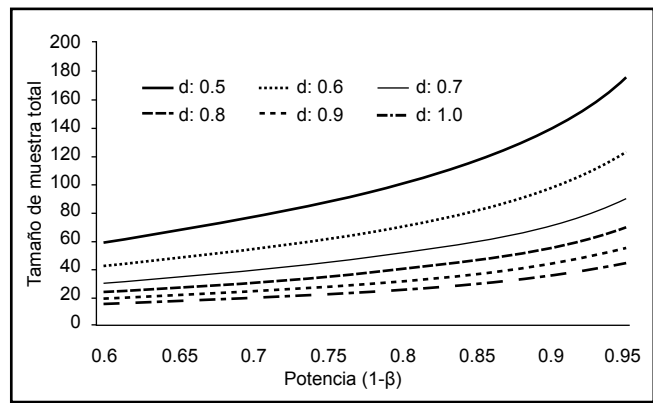


Figura 1. Relación entre tamaño de muestra, tamaño de efecto (d) y potencia, con un nivel de α de 0.05 al comparar los 2 promedios.

probabilidad de detectar diferencias si éstas realmente existen. Llevar a cabo estudios con baja potencia estadística no es éticamente aceptable, puede conducir a resultados de dudosa validez científica y a la subutilización de recursos de investigación.

Actualmente existen numerosos programas disponibles para calcular la potencia estadística y el tamaño de muestra apropiado para numerosos tipos de prueba. Uno de los más completos es el G-Power, que puede ser obtenido de forma gratuita del sitio: <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>.

Referencias

1. Cohen, J. 1992. A power premier. *Psychological Bulletin* 112: 155- 159.
2. Latour-Pérez, J. 2002. ¿Son éticos los ensayos clínicos con baja potencia estadística? *Revista electrónica de medicina intensiva. Debate* N° 4, Vol. 2, n° 8.
3. Lipsey, M. W. 1990. *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA. Sage.
4. Ramírez-Figueroa, C. y Sotres-Ramos, D. 2005. Tablas de potencia y tamaño de muestra de la prueba Uniformemente Más Potente Invariante (UMPI) para demostrar equivalencia de las medias de dos distribuciones normales. *Agrociencia* 39: 657-666.