

Estimación: intervalos de confianza

V. Abraira

Unidad de Bioestadística Clínica. Hospital Ramón y Cajal. Madrid.

Todas las investigaciones clínicas se realizan en un grupo determinado de individuos, generalmente pacientes. Sin embargo, el interés del estudio radica, habitualmente, en la generalización de los resultados: no se está tan interesado en lo que ocurre con los individuos particulares que participan en la investigación como en predecir lo que pueda ocurrir en el futuro con otros individuos similares. Esto plantea dos problemas, el primero es el del sesgo de selección¹, es decir, hasta qué punto los pacientes incluidos en el estudio son similares o, en la terminología estadística, pertenecen a la misma población que aquel o aquellos a quienes se quieren aplicar los resultados del estudio o, dicho de otra manera, hasta qué punto son tan distintos como para que los resultados del estudio no sean útiles para tomar las decisiones sobre éstos². El segundo problema es que los estudios nunca se realizan con todos los pacientes de interés, sino sólo con un grupo de ellos o, en terminología estadística, no se trabaja con la población sino con una muestra. La estadística sirve justamente para solucionar este segundo problema con las técnicas de estimación y contraste de hipótesis³.

Por ejemplo, en un estudio para evaluar la eficacia de los inhaladores de nicotina para reducir la cantidad de tabaco fumado⁴ se definió como éxito en cada individuo que, en el período comprendido entre la sexta semana y el cuarto mes desde el inicio del tratamiento, se redujera al menos en un 50% el número de cigarrillos fumados diariamente. Los participantes en el estudio fueron 400 voluntarios sanos, reclutados por anuncios en los periódicos, que estaban dispuestos a reducir su consumo de tabaco, pero no a dejar de fumar inmediatamente, o eran incapaces de ello. Los participantes fueron distribuidos aleatoriamente en dos grupos de 200, la evaluación de la eficacia se obtuvo de la comparación de las proporciones de éxito entre los grupos tratado (26%) y placebo (9%) y se pretende que este resultado sea aplicable a otros fumadores. En el artículo se describen las características demográficas de los participantes y las de su historia y hábito tabáquico para ayudar al lector a juzgar cuán diferentes puedan ser de otros fumadores a quienes se quiera aplicar el tratamiento (sesgo de selección). Una vez aceptado que son suficientemente parecidos, el siguiente

Puntos clave

- En general, el interés de los estudios de investigación clínica no es tanto lo que ocurre con los individuos particulares que participan en la investigación sino predecir lo que pueda ocurrir en el futuro con otros individuos similares.
- El intervalo de confianza de un índice es un intervalo dentro del que se encuentra el verdadero valor del índice con un grado prefijado de seguridad, suponiendo que la muestra sea aleatoria.
- El intervalo de confianza es, por tanto, un modo de cuantificar lo generalizable de los resultados de un estudio, suponiendo que el estudio es válido.

problema es cuánto tienen que ver las proporciones de éxito del 26 y el 9% obtenidas en los 200 individuos concretos que participan en cada grupo del estudio (muestra) con las proporciones respectivas en todos los individuos (población) a los que sea aplicable y cuánto dependen estas proporciones de que se obtengan en 200, 20 o 2.000 individuos.

Para ayudar a comprenderlo pensemos en otro problema similar: ¿cuál es la probabilidad de que al tirar un dado salga un dos? La diferencia entre ambos problemas reside en que en este último, si aceptamos que el dado está bien construido, podemos calcular que la probabilidad es 1/6, cálculo que no se puede realizar para el éxito de los inhaladores de nicotina. Sin embargo, también podríamos resolver el problema del dado tirándolo una serie de veces y calculando la proporción de veces en que sale el dos. En la figura 1 se ilustra la evolución de esta proporción en función del número de tiradas en 4 series. En ella se observa la característica forma de embudo correspondiente a una ley empírica que dice que la proporción oscila, pero que la amplitud de las oscilaciones va decreciendo a medida que aumenta el número de tiradas, de tal modo que en todas las series tiende a estabilizarse en el mismo valor, a pesar de que obviamente también es posible que en alguna serie salga el dos en todas las tiradas o no salga en ninguna. Se puede definir la probabilidad de un suceso co-

Correspondencia: Dr. V. Abraira.
Unidad de Bioestadística Clínica. Hospital Ramón y Cajal.
Ctra. Colmenar, km 9,100. 28034 Madrid.
Correo electrónico: victor.abraira@hrc.es

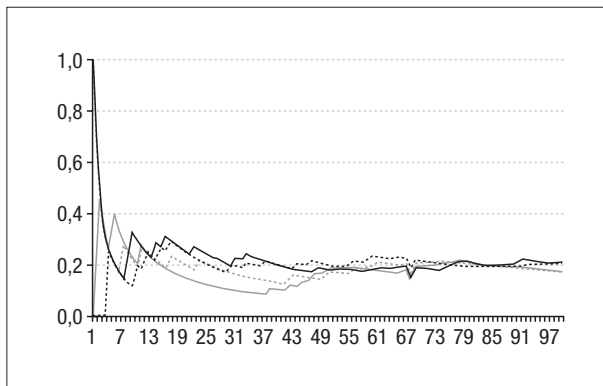


Figura 1. Evolución de la proporción del resultado "dos" en función del número de "tiradas" del dado en 4 series de 100 "tiradas" (los valores finales de las 4 series varían en este caso entre 0,17 y 0,21).

mo el valor en el que se estabiliza la proporción de veces en que el suceso ocurre. El problema consiste en saber, en cada caso, si el "número de tiradas" es suficientemente grande para que la proporción se haya estabilizado. Es decir, volviendo al estudio de los inhaladores, ¿el número de fumadores estudiados es suficiente para que las proporciones de éxito en cada grupo ya estén estabilizadas cerca de las verdaderas probabilidades de éxito? o, planteado de otra manera, dado el número de fumadores estudiados y las proporciones de éxito encontradas, ¿dentro de qué intervalo estarían las verdaderas probabilidades de éxito?

Imaginemos que no sabemos si el dado está bien construido o está "cargado", es decir, no sabemos si la probabilidad de que salga un dos es realmente $1/6$ ($= 0,166$). El experimento para estudiarlo consiste en tirar el dado varias veces. Supongamos que lo tiramos 6 veces y el dos sale en una ocasión. En este punto la mejor estimación de la probabilidad del dos es la proporción $1/6$, pero está claro

que, con tan sólo 6 tiradas, esta proporción puede estar muy alejada de la probabilidad y sería perfectamente posible encontrar este resultado con un dado cargado. En función de experiencias similares a la representada en la figura 1 se podría aceptar que la verdadera probabilidad podría estar en un amplio intervalo como, por ejemplo, 0,005 y 0,7. Piénsese, o constrúyase con un poco de paciencia, una gráfica como la de la figura 1 para el resultado "cara" en varias series de tiradas de una moneda. En la zona izquierda, las partes anchas de los embudos se solaparían. Si tiramos ahora el dado 24 veces y encontramos el dos en 4 ocasiones la mejor estimación sigue siendo $4/24 = 1/6$, pero el intervalo de probabilidades se estrecharía, digamos que entre 0,05 y 0,35. Si lo tiramos 120 veces y el dos se obtiene en 20 ocasiones, la estimación continúa siendo $20/120 = 1/6$, pero el intervalo se habría estrechado aún más, por ejemplo entre 0,1 y 0,2. Estos intervalos se denominan "intervalos de confianza" y las técnicas estadísticas permiten construirlos de una forma precisa y fiable, asumiendo que la muestra del estudio es una muestra aleatoria de la población de interés. En la tabla 1 se incluyen los intervalos para el problema del dado, contruidos con un nivel de confianza del 95 y el 99%.

Esta tabla ilustra cómo los intervalos de confianza nos indican si el tamaño de una muestra es suficiente para contestar a una pregunta de investigación. Si deseamos estar razonablemente seguros de que la "carga" del dado no altera la probabilidad del dos en más del 10% necesitamos menos de 120 tiradas, pero si queremos asegurar un error menor del 3% necesitaríamos 600.

En el caso de los inhaladores de nicotina los intervalos de confianza del 95% para el placebo son del 5,4-13,9%, y para el grupo tratado del 20,1-32,7%; que ambos no se solapan nos está indicando que el tamaño muestral es suficientemente grande para estar seguros, con un 95% de confianza, de que las proporciones de éxito son verdaderamente distintas en ambos grupos. Las técnicas estadísticas permiten construir intervalos de confianza no sólo para las proporciones sino también para todas las medidas del efecto de un tratamiento vistas en notas previas^{5,6}, y siempre debería darse el intervalo de confianza del índice usado. En la tabla 2 aparecen los distintos índices y sus intervalos de confianza para el ensayo de los inhaladores de nicotina.

Tabla 1. Intervalos de confianza para la proporción 1/6 en el experimento del dado

N.º de tiradas	N.º de dos	IC del 95%	IC del 99%
6	1	0,004-0,64	0,0008-0,74
12	2	0,02-0,48	0,009-0,57
24	4	0,05-0,37	0,03-0,44
48	8	0,07-0,30	0,06-0,35
120	20	0,11-0,25	0,09-0,27
600	100	0,14-0,20	0,13-0,21

Tabla 2. Medidas de eficacia y sus intervalos de confianza en el ensayo de los inhaladores de la nicotina

	Índice	IC del 95%
DAR	17,0	9,7-24,3
RR	2,89	1,75-4,76
DRR	188,9	108,2-269,5
OR	3,55	1,99-6,33
NNT	6	4-10

Como en este caso el riesgo del grupo experimental es mayor que el del grupo control, se ha preferido usar la diferencia de riesgo en lugar de la reducción de riesgo⁵. DAR: Diferencia absoluta de riesgo; RR: Riesgo relativo; DRR: Diferencia relativa de riesgo; OR: Odds ratio; NNT: Número necesario para tratar.

BIBLIOGRAFÍA

- Ellenberg JH. Selection bias in observational and experimental studies. Stat Med 1994;13:557-67.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Medicina basada en la evidencia. Cómo ejercer y enseñar la MBE. Madrid: Churchill Livingstone España, 1997.
- Abraira V. Estadística inferencial. Disponible en: <http://www.hrc.es/bioest/Introduccion.html> [consultado: 11/07/2001].
- Bolliger CT, Zellweger JP, Danielsson T, van Biljon X, Robidou A, Westin A, et al. Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety. BMJ 2000;321:329-33.
- Abraira V. Medidas del efecto de un tratamiento (I): reducción absoluta del riesgo, reducción relativa del riesgo y riesgo relativo. SEMERGEN 2000;26:535-6.
- Abraira V. Medidas del efecto de un tratamiento (II): odds ratio y número necesario para tratar. SEMERGEN 2001;27:418-20.