CONCEPTOS EN EPIDEMIOLOGIA

El significado de la significancia

Favio Rivas

Se puede afirmar que las pruebas de significancia estadística son métodos indirectos para evaluar los resultados de un estudio que no se puede hacer en toda la población-sujeto sino en una muestra de ella y que, fundamentalmente, buscan responder a la siguiente pregunta: ¿se hubiera encontrado lo mismo si en mi estudio estuviera toda la población?

Sin embargo, a las pruebas de significancia estadística se les ha atribuido más de lo que pueden dar. El problema hoy es mayor puesto que en la información científico-médica suministrada en la literatura o en los congresos y simposios -que por tanta es casi imposible de analizar a fondo- se termina dándole a p (*p value*) un alcance que no tiene y descartando información que puede ser muy valiosa desde el punto de vista clínico, epidemiológico o de salud pública, solo por resultar «no significativa estadísticamente».

En este artículo, la significancia estadística se aborda de manera sencilla con el ánimo de que quien acuda a la literatura científica o se exponga a los congresos, pueda evaluar -con un mínimo de criterio epidemiológico- muchos hallazgos que encontrará, en buena proporción, expresados con sentido estadístico.

Debe quedar claro, sin embargo, que la estadística es una de las herramientas con las cuales trabaja el epidemiólogo y con su ayuda se han iniciado muchas investigaciones que, partiendo del hallazgo (a veces accidental) de una asociación, terminaron en el control de una epidemia o en medidas de salud pública dirigidas a factores de riesgo específicos o en análisis de laboratorio, buscando micros-

cópicos causantes de enfermedad. El epidemiólogo necesita, sin lugar a dudas, valerse de esta herramienta para darle base firme a sus estudios, pero, requiere además del análisis juicioso de los hallazgos para darle sentido a sus conclusiones y razón a sus acciones en salud pública.

Mediante las pruebas de significancia estadística, se pretende -en últimas- establecer asociaciones, las cuales se fundamentan en la detección de diferencias. Por ejemplo, un estudio cualquiera encontró diferencia en la ocurrencia de determinado proceso patológico entre una población expuesta a un factor de riesgo y ello permite pensar que la exposición está asociada con la ocurrencia de la enfermedad. Pero, claro, tal diferencia pudo haber ocurrido por simple casualidad; es decir, puede aparecer inesperadamente por efecto del azar y no necesariamente porque exista una real diferencia. Como se repetirá más adelante, lo que quiere indagarse, entonces, es si tal diferencia (que conlleva una asociación) detectada en el estudio pueda haber ocurrido por azar.

Supuestos erróneos acerca del valor de p

- Un valor bajo de p (por ejemplo, $p=0,0002$) prueba que la diferencia encontrada se debe a la intervención, si es un estudio experimental, o a la exposición en uno de cohortes o de casos y controles.
- Un valor bajo de p demuestra relación causa-efecto entre una exposición y un padecimiento.
- Un valor bajo de p establece la fuerza de la asociación.

- Un valor bajo de p demuestra la validez del estudio.

- Un valor de p alto (por ejemplo, $p > 0,05$) significa que no hay asociación entre la exposición A y el resultado B o que no hay diferencia entre exposición y no exposición, etc.

Desde luego, hay algunas otras interpretaciones erróneas con respecto al valor de p , pero quizás, éstas sean las más frecuentemente leídas, referidas o escuchadas por parte del personal de salud.

Las pruebas de significancia estadística parten de una hipótesis de no asociación (o de no diferencia) a la cual se le llama hipótesis nula. Por ejemplo, se quiere evaluar mediante un ensayo clínico, si el tratamiento B es mejor que el ya conocido tratamiento A para el manejo de la hipertensión arterial. La hipótesis nula sería que el tratamiento B es igual (o no es mejor) que el A, a la cual se le contrapone una hipótesis alterna: el tratamiento B es mejor que el tratamiento A. La diferencia de proporciones de individuos que mejoraron con uno y otro tratamiento, se evaluará con la prueba de significancia.

Vamos a suponer que una vez finalizado el estudio, se encontró una diferencia (estadísticamente significativa) en las proporciones de individuos que mejoraron en cada grupo de pacientes hipertensos, a favor del tratamiento B. La prueba de significancia estadística solamente permitirá establecer qué tan probable era encontrar esa diferencia (o una mayor) de ser cierta la hipótesis nula. Es decir, si partimos del supuesto que los dos tratamientos contra la hipertensión arterial eran similares, ¿qué tan probable era encontrar la diferencia que pudimos detectar o una aún mayor a esa? Esta probabilidad es la que está indicando el valor de p . Si resultó ser del orden de 0,002, quiere decir que la probabilidad de detectar tal diferencia (o una más grande) entre las proporciones de individuos hipertensos que mejoraron con uno y otro tratamiento, es de 2 por mil y eso se considera ínfimo. Por tanto, se afirma que los

dos tratamientos no tienen un efecto similar sobre la hipertensión arterial (rechazo de la hipótesis nula) y, por extensión, que el medicamento B es mejor que el A.

El primer concepto que debe quedar claro, entonces, es que las pruebas estadísticas nada prueban; simplemente evalúan la probabilidad de que nuestros hallazgos puedan o no atribuirse al azar. Para probar una asociación en el campo de la ciencia, no son suficientes las pruebas estadísticas y se necesitan varios criterios adicionales con los cuales puede establecerse, por ejemplo, una relación causa-efecto, como se verá más adelante.

En contraste con la probabilidad de que ocurra por azar una diferencia detectada en un estudio (valor p), las medidas de asociación (riesgo relativo, RR, y razón de disparidades, OR) indican la dirección y la fuerza con que están ligados dos factores (por ejemplo, exposición-enfermedad) y sus intervalos de confianza del 95% (IC95%) pueden ser usados como prueba de significancia estadística.

El RR se obtiene dividiendo la incidencia de enfermedad de quienes estaban expuestos al factor de riesgo entre la incidencia de enfermedad de quienes no lo estaban. Si las dos incidencias resultaren iguales, el RR será igual a 1 y a esto se llama el valor nulo de esta medida. El OR mide lo mismo que el RR en aquellos estudios en los cuales no es posible establecer las incidencias de la enfermedad (casos y controles).

Por su parte, los intervalos de confianza del 95% (IC 95%), para estos casos, indican entre qué valores se encuentra un RR o un OR con un 95% de certeza, es decir, con sólo 5% de incertidumbre. Aquellos IC 95% que dejan por fuera el valor nulo (uno) de cualquiera de estas dos medidas, serán significativos. Por ejemplo, se encontró un RR de 2 con un IC 95% que oscila entre 1,5 y 3,0; entonces, decimos que el hallazgo es estadísticamente significativo porque el valor nulo de todo RR es 1 y, en este caso, está por fuera del intervalo encontrado.

Sin embargo, como cualquier otra prueba de significancia, también los intervalos tienen algo que ver con el tamaño de la muestra. Cuando es pequeña, los IC 95% serán muy amplios y pueden llegar a incluir la unidad, lo cual -estadísticamente hablando- haría tal hallazgo «no significativo» y posiblemente, si se hubiera obtenido una muestra de mayor tamaño, se podría encontrar que los IC 95% no incluyen la unidad, esto sería «estadísticamente significativos».

Un ejemplo de lo anterior se muestra en un estudio realizado hace algunos años, el cual buscaba la asociación entre la presencia de úlceras genitales y el desarrollo de seroconversión al VIH (figura 1).

Figura 1. Seguimiento de pacientes con presencia de úlceras genitales y sin ellas con respecto al desarrollo de positividad al VIH.

		Seroconversión al VIH			
		Sí	No		
Úlceras genitales	Sí	8	45	53	RR=3,55 IC 95% (0,8-15,9) p=0,07
	No	2	45	47	
		10	90	100	

		Seroconversión al VIH			
		Sí	No		
Úlceras genitales	Sí	16	90	106	RR= 3,55 IC95% (1,2-10,2) p= 0,02
	No	4	90	94	
		20	180	200	

Se observa cómo, teóricamente, al estudiar una muestra mayor, ese mismo RR adquiere unos intervalos de confianza menos amplios, que dejan por fuera la unidad, y el valor de p, a su vez, se hace más pequeño, es decir, se pasa desde la no significancia a la significancia estadística.

En resumen, con un valor de p de 0,05, lo único que se quiere decir es que de 20 estudios semejantes en los cuales se encontró una

asociación estadísticamente significativa, en uno de ellos tal asociación pudo ser producto del azar; o dicho de otra manera, de 100 estudios similares, en cinco, tal diferencia o asociación encontrada, ocurrió accidentalmente y no porque existiera en realidad.

El segundo concepto es que el no poder rechazar la hipótesis nula, nunca significa que la alterna sea falsa. O sea, que un valor de p mayor de 0,05 (5%) no necesariamente significa la inexistencia de asociación (o de diferencia entre los dos tratamientos a favor del B, en el ejemplo). Ese dato solamente informa que la diferencia encontrada en el estudio pudo ocurrir por azar. En otras palabras, que no podemos descartar el papel del azar en nuestro hallazgo.

Es claro que el papel del azar nunca va a poder descartarse en el 100%, por más pequeño que resulte el valor de p. Eso, precisamente, hace la diferencia entre lo improbable y lo imposible; un valor de p muy pequeño hace improbable el hecho de que una diferencia (o asociación) igual o más extrema que la detectada en el estudio, pueda ser explicada por el azar, pero no la hace imposible.

Si un determinado estudio en una muestra pequeña de la población-sujeto resulta con un riesgo relativo alto (mayor de 2) en ausencia de sesgos o de factores de confusión, aún con unos intervalos de confianza que incluyan la unidad, debe ser tenido en cuenta por cuanto, a pesar de la no significancia estadística, puede ser epidemiológicamente relevante. Un claro ejemplo (y extremo, por demás) se observa en un estudio de cohorte con 215 VIH positivos y 201 VIH negativos para establecer el riesgo de desarrollar TBC clínica. Se encontró un RR de 7,5 con unos IC 95% de 0,9 a 59, que resultó no significativo desde el punto de vista estadístico (por estar incluido el valor nulo del RR en ese intervalo). Pero, estaríamos completamente ciegos si no le diéramos importancia epidemiológica a semejante asociación (figura 2).

De lo anterior, se puede desprender que la fuerza de la asociación no puede determinarse

Figura 2. Seguimiento de pacientes VIH positivos y negativos con relación al desarrollo de TBC clínica, todos PPD positivos.

		TBC clínica		
		Sí	No	
Úlceras genitales	Sí	32	180	212
	No	8	180	188
		40	360	400

RR=3,55
IC 95% (1,7-7,2)
p=0,005

con el valor de P pero si con un RR o un OR, por cuanto estos datos (en el caso más sencillo) dicen cuántas veces es mayor el riesgo de enfermar entre quienes se expusieron con respecto a quienes no lo hicieron. Si el RR es 3, entonces, los expuestos tienen tres veces mayor probabilidad de enfermar que quienes no se expusieron. Los RR y los OR, con el apoyo de los intervalos de confianza, nos muestran también la dirección de la asociación. En el caso del referido RR de 3, se ha dicho que quienes están expuestos tienen un riesgo mayor de enfermar, pero, puede ocurrir que el RR sea menor de 1 y que los intervalos de confianza del 95% también dejen por fuera el valor nulo. En este caso, deberemos decir que quienes se exponen están protegidos de la enfermedad o que la exposición no es un factor de riesgo y sí uno de protección. Esta es la dirección de la asociación, algo que no puede obtenerse con el valor de p.

Adicionalmente, los RR u OR mayores de 1 pero menores de 2, que a simple vista pudieran estimarse bajos, sin importancia desde el punto de vista de la fuerza de la asociación, pueden ser extraordinariamente importantes si se miran desde la óptica de la salud pública. Cuando hay una alta prevalencia de la exposición o del factor de riesgo en la comunidad, los beneficios que se obtendrían, si se controlara dicha exposición, pudieran ser valiosos. Esto último tiene que ver con otra medida en epidemiología, relacionada con la frecuencia con que se presenta una característica en la población, el riesgo poblacional atribuible. Esto es, la proporción máxima de incidencia de una enfermedad que lograría eliminarse si se controlara completamente la exposición implicada.

El cuadro adjunto (cuadro 1), extraído del texto de Lilienfeld, muestra cómo, con un RR de 2 y una prevalencia de la característica susceptible de ser controlada totalmente de 95%, la incidencia de tal enfermedad se podría disminuir en 49%.

Cuadro 1. Riesgos atribuibles, expresados en términos de proporción, correspondientes a un cierto valor de riesgo relativo y a una cierta proporción de la población con determinada característica.

Porcentaje de la población con característica	Riesgo relativo			
	2	4	10	12
10	0,09	0,23	0,47	0,52
30	0,23	0,47	0,73	0,77
50	0,33	0,60	0,82	0,84
70	0,41	0,67	0,86	0,89
90	0,47	0,73	0,89	0,91
95	0,49	0,74	0,90	0,92

Finalmente, se pueden encontrar asociaciones estadísticamente significativas que no tienen ninguna significancia desde puntos de vista no estrictamente estadísticos. Muestras de estudio en extremo grandes, pueden dejar ver diferencias significativas que, en realidad, son irrelevantes clínica o epidemiológicamente hablando.

En resumen, el asunto de la significancia estadística es un hallazgo estadístico que en ciertas ocasiones, cuando la muestra del estudio es adecuada y sin sesgos o factores de confusión, puede ser de valor clínico, epidemiológico o de salud pública.

El tercer concepto que debe quedar claro, es que la significancia estadística por sí misma no implica relación causa-efecto y tampoco la no significancia la descarta. Aparte del papel que haya podido estar jugando el azar que, a su vez, tiene bastante que ver con el tamaño de la muestra, se deben evaluar otros aspectos. Uno de tales aspectos es la presencia o no de sesgos en el diseño del estudio que, fundamentalmente, pueden ser de selección, de información o de confusión, además de los errores cometidos en el análisis de sus resultados. No es extraño utilizar una prueba

estadística inadecuada para el tipo de datos que se está analizando, aunque ahora, con la ayuda de los computadores y de los paquetes estadísticos es menos probable encontrar estas fallas en la literatura científica.

Los sesgos, que resultan críticos en cualquier tipo de estudio, determinan la validez y, si a su prevención, le adicionamos el juicio del epidemiólogo al considerarlos, así como la consistencia de los hallazgos con anteriores estudios similares, la relación temporal entre exposición y enfermedad, la plausibilidad biológica y la relación dosis-respuesta, puede considerarse, ahora sí con más seriedad, una relación causa-efecto.

No se puede finalizar esta revisión sin mencionar un último punto, para tranquilidad de los estadísticos y, también, de los epidemiólogos. Si se tiene un estudio bien hecho, esto es, con una muestra adecuada para el problema que desea abordarse, libre de sesgos y de confusión y con una aplicación correcta de las pruebas de significancia, insistir en que hay asociación sólo porque se observó cierta diferencia en los resultados (por ejemplo, de los tratamientos A y B mencionados anteriormente) sin que se consiguiera rechazar la hipótesis nula, es a todas luces, inadecuado.

De otro lado, los estudios que muestran asociaciones estadísticas significativas (cuando se describen por primera vez) son y han sido, en no pocas ocasiones, el punto de partida de investigaciones más profundas que han permitido conocer a fondo algunos factores de riesgo de enfermedad.

Finalmente, K. Rothman y S. Greenland, en la última edición de su texto *Modern epidemiology*, afirman "ser estadístico o médico o ambos al tiempo, no es necesario ni suficiente para ser epidemiólogo. Se necesita el entendimiento de los principios de la investigación epidemiológica y experiencia para aplicarlos". De eso se trata, de pensar con criterio epidemiológico, cosa bastante diferente del simplemente estadístico o del simplemente médico. Es mucho más «fácil»

atribuir al azar un hallazgo en contra de una asociación ya descrita y estudiada que cuando es la primera vez que se describe, independientemente del resultado de las pruebas.

Si con lo expuesto hasta ahora no ha quedado clara la respuesta a la pregunta de qué tratan de resolver las pruebas de significancia, aquella con que empieza este documento, hay que decir que no es posible responderla ciento por ciento afirmativamente y a lo máximo que podremos llegar es al 90%, al 95% o al 99% de certidumbre (este último con tamaños de muestra bastante mayores que los dos primeros porcentajes). Es decir, si con nuestro estudio (eso sí, bien hecho) hemos detectado un RR de 2,2 con unos intervalos de confianza del 95% entre 1,7 y 3,1, podemos afirmar con un 95% de probabilidades de no equivocarnos, que el verdadero RR en el universo poblacional estará entre 1,7 y 3,1. A esto, nunca podremos llegar al considerar únicamente el valor de p.

Referencias

- Beaglehole R, Bonita R, Kjellstrom T.** Epidemiología básica. Publicación Científica No. 551. Washington, D.C.: OPS; 1994.
- Giesecke J.** Modern infectious disease epidemiology. New York: Oxford University Press, Inc.; 1994.
- Gregg M.** Field epidemiology. New York: Oxford University Press, Inc.; 1996.
- Norman G, Streiner D.** Bioestadística. Barcelona: Mosby/Doyma Libros; 1996.
- Londoño J.** Metodología de la investigación epidemiológica. Medellín: Editorial Universidad de Antioquia; 1995.
- Dawson B, Trapp R.** Basic and clinical bio-statistics. New Jersey: Prentice-Hall International Inc.; 1990.
- Goldberg M.** La epidemiología sin esfuerzo. Madrid: Diaz de Santos; 1994.
- Lilienfeld A, Lilienfeld D.** Fundamentos de epidemiología. México, D.F., Addison-Wesley Iberoamérica; 1986.
- Rothman K, Greenland S.** Modern epidemiology. Second edition. USA: Lippincott-Raven Publishers; 1998.