

Cálculo de probabilidades: nociones básicas

Pértega Díaz, S.¹; Pita Fernández, S.²

¹Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Universitario Juan Canalejo (A Coruña). ²Centro de Salud de Cambre (A Coruña).

CAD. ATEN. PRIMARIA 2004; 11: 176-179

La estadística, junto con la epidemiología, es un instrumento indispensable en el proceso de investigación en medicina. Formalmente, se puede clasificar la estadística en **descriptiva**, cuando se utiliza simplemente para la presentación y síntesis de la información recogida en un estudio, e **inferencial**, que tiene por objetivo generalizar la información obtenida en una muestra a resultados válidos para la población de la que procede¹. Supongamos, por ejemplo, que nos interesa comparar dos fármacos A y B y determinar cuál de ellos es más eficaz para el tratamiento de una determinada enfermedad. Para ello, se diseña un estudio distribuyendo 100 enfermos en dos grupos, cada uno de los cuales recibe uno de los dos tratamientos. Al cabo de 1 mes, la tasa de curación en cada grupo es del 80% y del 70%, respectivamente. Ante esta información, ¿es correcto suponer que el tratamiento A es mejor que el tratamiento B para esta enfermedad en concreto? La respuesta a esta pregunta, como a la mayor parte de problemas que pueden plantearse en medicina, está sujeta a un cierto grado de incertidumbre que hacen muy complicado tomar una decisión al respecto. En la respuesta de un paciente al tratamiento pueden influir diversos factores, entre los que se incluye el azar, que pueden provocar una gran variabilidad en los resultados. La aplicación de los principios de la estadística a la clínica permite reducir y cuantificar dicha variabilidad y ayudar a la toma de decisiones. En particular, el **cálculo de probabilidades** suministra las reglas apropiadas para cuantificar esa incertidumbre y constituye la base para la estadística inductiva o inferencial.

El objetivo de este trabajo consiste en introducir algunos de los conceptos básicos del cálculo de probabilidades, así como las reglas necesarias para el desarrollo de la inferencia estadística en medicina. Una exposición más detallada de estos y otros conceptos puede encontrarse en referencias más especializadas²⁻⁸.

El concepto de **probabilidad** resulta familiar a cualquier profesional del ámbito sanitario, pero una definición más precisa exige considerar la naturaleza matemática de dicho concepto. La probabilidad de ocurrencia de un determinado suceso podría definirse como la proporción de veces que ocurriría dicho suceso si se repitiese un experi-

mento o una observación en un número grande de ocasiones bajo condiciones similares. Por definición, entonces, la probabilidad se mide por un número entre cero y uno: si un suceso no ocurre nunca, su probabilidad asociada es cero, mientras que si ocurriese siempre su probabilidad sería igual a uno. Así, las probabilidades suelen venir expresadas como decimales, fracciones o porcentajes.

La definición anterior de probabilidad corresponde a la conocida como definición frecuentista. Existe otra descripción más formal desde el punto teórico que permite definir el concepto de probabilidad mediante la verificación de ciertos axiomas a partir de los que se deducen todas las demás propiedades del cálculo de probabilidades². En otros contextos, se ha defendido una interpretación más amplia del concepto de probabilidad que incluye las que podemos denominar probabilidades subjetivas o personales, mediante las cuales se expresa el grado de confianza o experiencia en una proposición. Esta definición constituye la base de los llamados métodos bayesianos, que se presentan como alternativa a la estadística tradicional centrada en el contraste de hipótesis⁹⁻¹¹. No obstante, y en relación con el propósito de este trabajo, bastará con considerar la definición frecuentista anterior. Así, a partir de una población con N elementos, de los cuales k presentan una característica A, se estimará la probabilidad de la característica A como $P(A) = k/N$. Así, por ejemplo, en una población de 100 pacientes, 5 de los cuales son diabéticos, la probabilidad de padecer diabetes $p(\text{Diabetes})$ se estimará como el cociente: $e 5/100 = 0.5$.

Es conveniente conocer algunas de las propiedades básicas del cálculo de probabilidades:

- Para un suceso A, la probabilidad de que suceda su complementario (o equivalentemente, de que no suceda A) es igual a uno menos la probabilidad de A:

$$P(A) + P(\bar{A}) = 1 \Rightarrow P(\bar{A}) = 1 - P(A)$$

donde A denota al suceso contrario o suceso complementario de A.

- Si un fenómeno determinado tiene dos posibles resultados A y B mutuamente excluyentes (es decir, que no pue-

den darse de forma simultánea, como ocurre en el lanzamiento de una moneda al aire), la probabilidad de que una de esas dos posibilidades ocurra se calcula como la suma de las dos probabilidades individuales:

$$P(A \text{ ó } B) = P(A) + P(B) \quad (1)$$

La extensión de la ley aditiva anterior al caso de más de dos sucesos mutuamente excluyentes A, B, C... indica que:

$$P(A \text{ ó } B \text{ ó } C \text{ ó } \dots) = P(A) + P(B) + P(C) + \dots$$

Consideremos, como ejemplo, un servicio de urología en el que el 38,2% de los pacientes a los que se les practica una biopsia prostática presentan una hiperplasia benigna (HB), el 18,2% prostatitis (PR) y en un 43,6% el diagnóstico es de cáncer (C). La probabilidad de que en un paciente que se somete a una biopsia de próstata no se confirme el diagnóstico de cáncer prostático será igual a:

$$P(HB \text{ ó } PR) = P(HB) + P(PR) = 0,382 + 0,182 = 0,564$$

Es decir, en un 56,4% de los casos se logra descartar un diagnóstico maligno. De modo equivalente, la probabilidad anterior podría haberse calculado como la probabilidad del suceso contrario al del diagnóstico de cáncer:

$$P(HB \text{ ó } PR) = P(\bar{C}) = 1 - P(C) = 1 - 0,436 = 0,564''$$

Nótese la importancia del hecho de que los sucesos anteriores sean mutuamente excluyentes. Sin esta condición, la ley de adición no será válida. Por ejemplo, se sabe que en una determinada Unidad de Cuidados Intensivos (UCI) el 6,9% de los pacientes que ingresan lo hacen con una infección adquirida en el exterior, mientras que el 13,7% adquieren una infección durante su estancia en el hospital. Se conoce además que el 1,5% de los enfermos ingresados en dicha unidad presentan una infección de ambos tipos. ¿Cuál será entonces la probabilidad de que un determinado paciente presente una infección de cualquier tipo en UCI? Para realizar el cálculo, si se suman simplemente las probabilidades individuales (0,069+0,137) la probabilidad de un suceso doble (infección comunitaria y nosocomial) se estará evaluando dos veces, la primera como parte de la probabilidad de padecer una infección comunitaria y la segunda como parte de la probabilidad de adquirir una infección en la UCI. Para obtener la respuesta correcta se debe restar la probabilidad del doble suceso. Así:

- Si un fenómeno determinado tiene dos posibles resultados A y B, la probabilidad de que una de esas dos

posibilidades ocurra viene dada, en general, por la expresión:

$$P(A \text{ ó } B) = P(A) + P(B) - P(A \text{ y } B)$$

Por lo tanto, si dos o más sucesos no son mutuamente excluyentes, la probabilidad de que ocurra uno de ellos o ambos se calcula sumando las probabilidades individuales de que ocurra una de esas circunstancias, pero restando la probabilidad de que ocurra la común.

Resulta evidente que, para el caso de procesos mutuamente excluyentes, $P(A \text{ y } B)$ y se obtiene (1).

En el ejemplo anterior, la probabilidad de infección en UCI vendrá dada, por lo tanto, como:

$$P(A \text{ ó } B) = 0,069 + 0,137 - 0,015 = 0,191$$

Es decir, 19 de cada 100 enfermos registrará alguna infección (ya sea de tipo comunitario o nosocomial) durante su ingreso en la citada unidad.

A veces, la probabilidad de que un determinado suceso tenga lugar depende de que otro suceso se haya producido o no con anterioridad. Esto es, en ocasiones el hecho de que se produzca un determinado fenómeno puede hacer más o menos probable la aparición de otro. Este tipo de probabilidades se denominan **probabilidades condicionadas**, y se denotará por $P(A/B)$ a la probabilidad condicionada del suceso A suponiendo que el suceso B haya ocurrido ya.

- La ley **multiplicativa de probabilidades** indica que la probabilidad de que dos sucesos A y B ocurran simultáneamente es igual a:

$$P(A \text{ y } B) = P(A/B) \cdot P(B) \quad (3)$$

La ley multiplicativa anterior se utiliza también con el fin de determinar una probabilidad condicional $P(A/B)$ a partir de los valores de $P(A \text{ y } B)$ y $P(B)$:

$$P(A/B) = \frac{P(A \text{ y } B)}{P(B)} \quad (4)$$

Supongamos, por ejemplo, que queremos estudiar la influencia del hecho de ser fumador como factor de riesgo en el desarrollo de una enfermedad en una determinada población. Para ello se diseñó un estudio prospectivo y, tras seleccionar una muestra de 180 sujetos, los resultados son los que se muestran en la Tabla 1. Considerando toda la muestra, la probabilidad de desarrollar la enfermedad (E) en la población de estudio es:

$$P(E) = \frac{80}{180} = 0,444 \Rightarrow 44,4\%$$

Mientras que la probabilidad de padecer la enfermedad un fumador (F) es:

$$P(E/F) = \frac{60}{70} = 0,857 \Rightarrow 85,7\%$$

Y un no fumador:

$$P(E/\bar{F}) = \frac{20}{110} = 0,182 \Rightarrow 18,2\%$$

Teniendo en cuenta que:

$$P(F) = \frac{70}{180} = 0,389 \Rightarrow 38,9\% \quad P(\bar{F}) = 1 - P(F) = 1 - 0,389 = 0,611 \Rightarrow 61,1\%$$

$$P(F \text{ y } E) = \frac{60}{180} = 0,333 \Rightarrow 33,3\% \quad P(\bar{F} \text{ y } E) = \frac{20}{180} = 0,111 \Rightarrow 11,1\%$$

Podría haberse aplicado la fórmula (4) para obtener cualquiera de las dos probabilidades condicionadas anteriores, resultando idénticos valores:

$$P(E/F) = \frac{P(E \text{ y } F)}{P(F)} = \frac{0,333}{0,389} = 0,857 \Rightarrow 85,7\% \quad P(E/\bar{F}) = \frac{P(E \text{ y } \bar{F})}{P(\bar{F})} = \frac{0,111}{0,611} = 0,182 \Rightarrow 18,2\%$$

En el ejemplo, se constata por lo tanto que la incidencia de la enfermedad es diferente en la población fumadora que en la no fumadora (85,7% vs 18,2%). Así pues, la probabilidad de desarrollar la enfermedad depende de si se es o no fumador. En otras ocasiones, sin embargo, sucede que la ocurrencia o no de un determinado fenómeno B no influye en la ocurrencia de otro suceso A. Se dice entonces que los sucesos A y B son **independientes** y se verificará que:

$$P(A/B) = P(A) \quad (5)$$

Sustituyendo (5) en (3) se obtiene entonces que:

$$P(A \text{ y } B) = P(A) \cdot P(B)$$

Es decir, en caso de independencia, la probabilidad de que ocurran dos sucesos de forma simultánea es igual al producto de las probabilidades individuales de ambos sucesos. Así, dos sucesos son independientes, si el resultado de uno no tiene efecto en el otro; o si el que ocurra el primero de ellos no hace variar la probabilidad de que se de el segundo. Obviamente, en la práctica, y debido a las variaciones en el muestreo, será extremadamente difícil encontrar una muestra que reproduzca de forma exacta las condiciones de independencia anteriores. El determinar si las diferencias observadas son o no compatibles con la hipótesis de independencia constituye uno de los principales problemas que aborda la estadística inferencial.

- Si se considera un fenómeno con k resultados posibles, mutuamente excluyentes, B_1, B_2, \dots, B_k y se conoce la pro-

babilidad de cada uno de ellos, el llamado **Teorema de las Probabilidades Totales** permite calcular la probabilidad de un suceso A a partir de las probabilidades condicionadas:

$$P(A) = P(A \text{ y } B_1) + P(A \text{ y } B_2) + \dots + P(A \text{ y } B_k)$$

Utilizando la expresión para el cálculo de la probabilidad de la intersección de dos sucesos se tiene que:

$$P(A \text{ y } B_k) = P\left(\frac{A}{B_k}\right) \cdot P(B_k)$$

y, por lo tanto:

$$P(A) = P(A/B_1) \cdot P(B_1) + P(A/B_2) \cdot P(B_2) + \dots + P(A/B_k) \cdot P(B_k) \quad (6)$$

En el ejemplo anterior, podría aplicarse este resultado para el cálculo de la incidencia de la enfermedad en la población de estudio:

$$P(E) = P(E/F) \cdot P(F) + P(E/\bar{F}) \cdot P(\bar{F}) = 0,857 \times 0,389 + 0,182 \times 0,611 = 0,444 \Rightarrow 44\%$$

Las leyes aditiva y multiplicativa, junto con la noción de probabilidades condicionadas y el teorema de las probabilidades totales se han empleado para desarrollar el llamado Teorema de Bayes, de indudable interés en la aplicación de la estadística al campo de la medicina. Si se parte de la definición de probabilidad condicionada (4):

$$P(A/B) = \frac{P(A \text{ y } B)}{P(B)} \quad \text{ó} \quad P(B/A) = \frac{P(A \text{ y } B)}{P(A)} \Rightarrow P(A \text{ y } B) = P(B/A) \cdot P(A)$$

siempre que $P(A) \neq 0$ y $P(B) \neq 0$. Aplicando además el teorema de las probabilidades totales se llega a que:

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B/A) \cdot P(A) + P(B/\bar{A}) \cdot P(\bar{A})}$$

El diagnóstico médico constituye un problema típico de aplicación del Teorema de Bayes en el campo médico, puesto que permite el cálculo de la probabilidad de que un paciente padezca una determinada enfermedad una vez dados unos síntomas concretos. La capacidad predictiva de un test o de una prueba diagnóstica suele venir dada en términos de su sensibilidad y especificidad¹². Tanto la sensibilidad como la especificidad son propiedades intrínsecas a la prueba diagnóstica, y definen su validez independientemente de cuál sea la prevalencia de la enfermedad en la población a la cual se aplica. Sin embargo, carecen de utilidad en la práctica clínica, ya que sólo proporcionan información acerca de la probabilidad de obtener un resultado concreto (positivo o negativo) en función de si un paciente

TABLA 1

Asociación entre el hábito tabáquico y el desarrollo de una enfermedad. Datos de un estudio de seguimiento en 180 individuos.

	Enfermos	Sanos	
Fumador	60	10	70
No fumador	20	90	110
	80	100	180

está realmente enfermo o no. Por el contrario, el concepto de valores predictivos, a pesar de ser de enorme utilidad a la hora de tomar decisiones clínicas y transmitir información sobre el diagnóstico, presenta la limitación de que dependen en gran medida de lo frecuente que sea la enfermedad a diagnosticar en la población objeto de estudio. El Teorema de Bayes permite obtener el valor predictivo asociado a un test al aplicarlo en poblaciones con índices de prevalencia muy diferentes.

Consideremos como ejemplo un caso clínico en el que una gestante se somete a la prueba de sobrecarga oral con 50 gramos de glucosa para explorar la presencia de diabetes gestacional, obteniéndose un resultado positivo. Es sabido que dicho test presenta unos valores aproximados de sensibilidad y especificidad en torno al 80% y al 87%, respectivamente. Si se conoce además que la prevalencia de diabetes gestacional en la población de procedencia es aproximadamente de un 3%, por medio del teorema de Bayes podemos conocer la probabilidad de que el diagnóstico sea correcto o, equivalentemente, el valor predictivo positivo:

$$P(D/+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|\bar{D}) \cdot P(\bar{D})} = \frac{\text{Sensibilidad} \times \text{Prevalencia}}{\text{Sensibilidad} \times \text{Prevalencia} + (1 - \text{Especificidad}) \times (1 - \text{Prevalencia})}$$

$$= \frac{0,80 \cdot 0,03}{0,80 \cdot 0,03 + (1 - 0,87) \cdot (1 - 0,03)} = \frac{0,024}{0,024 + 0,1261} = 0,159 \Rightarrow 15,9\%$$

Se puede concluir por lo tanto que, a pesar de obtener un resultado positivo en la prueba, existe sólo una probabilidad de un 15,9% de que la paciente padezca diabetes gestacional.

Supongamos que además dicha paciente tiene más de 40 años de edad. Se sabe que en grupos de edad más avanzada la prevalencia de diabetes gestacional entre las gestantes llega a aumentar hasta aproximadamente un 8%. En este caso, el valor predictivo positivo asociado vendrá dado por:

$$P(D/+) = \frac{\text{Sensibilidad} \times \text{Prevalencia}}{\text{Sensibilidad} \times \text{Prevalencia} + (1 - \text{Especificidad}) \times (1 - \text{Prevalencia})}$$

$$= \frac{0,80 \cdot 0,08}{0,80 \cdot 0,08 + (1 - 0,87) \cdot (1 - 0,08)} = \frac{0,064}{0,064 + 0,1196} = 0,3486 \Rightarrow 34,86\%$$

En este caso las posibilidades de un diagnóstico de diabetes gestacional aumentan hasta un 34,86%.

En un caso como este, en que se realiza una prueba para obtener información sobre un diagnóstico, suele hablarse de probabilidad a priori, que es la disponible antes de realizar la prueba (la prevalencia, en este caso) y probabilidad a posteriori, que es la obtenida después de realizarla (los valores predictivos). A su vez, se suele denominar verosimilitudes a las probabilidades de un suceso bajo distintas hipótesis. El teorema de Bayes permite así obtener los valores de las probabilidades a posteriori a partir de las probabilidades a priori mediante una multiplicación proporcional a las verosimilitudes.

Tal y como se indicó al inicio del presente artículo, la teoría de la probabilidad constituye la base matemática para la aplicación de la estadística inferencial en medicina. El cálculo de probabilidades constituye una herramienta que permitirá hacer inferencia sobre distintos parámetros poblacionales a partir de los resultados obtenidos en una muestra, y después tomar decisiones con el mínimo riesgo de equivocación en situaciones de incertidumbre.

BIBLIOGRAFÍA

- 1- Argimón Pallás JM, Jiménez Villa J. Métodos de investigación clínica y epidemiológica. 2ª ed. Madrid: Harcourt; 2000.
- 2- Senfís J, Pardell H, Cobo E, Canela J. Bioestadística. 3ª ed. Barcelona: Masson; 2003.
- 3- Colton T. Estadística en medicina. Barcelona: Salvat; 1979.
- 4- Armitage P, Berry G. Estadística para la investigación biomédica. Barcelona: Doyma; 1992.
- 5- Departamento de Medicina y Psiquiatría. Universidad de Alicante. Tratado de Epidemiología Clínica. Madrid: DuPont Pharma; 1995.
- 6- Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall; 1991.
- 7- Vélez R, Hernández V. Cálculo de Probabilidades I. Madrid: UNED; 1995.
- 8- Quesada V, García A. Lecciones de Cálculo de Probabilidades. Madrid: Díaz de Santos; 1988.
- 9- Silva LC, Benavides A. El enfoque bayesiano: otra manera de inferir. Gac Sanit 2001; 15(4): 341-346.
- 10- Silva LC, Suárez P. ¿Qué es la inferencia bayesiana? JANO 2000; 58: 65-66.
- 11- Silva LC, Muñoz A. Debate sobre métodos frecuentistas vs bayesianos. Gac Sanit 2000; 14: 482-494.
- 12- Pértega Díaz S, Pita Fernández S. Pruebas diagnósticas. Cad Aten Primaria 2003; 10: 120-124.