

Capítulo 3:

Probabilidad

Presentación

No es ningún secreto descubrir que muchos fenómenos son inciertos. A las mentes más deterministas les gusta pensar que esta incertidumbre es el resultado de nuestra falta de conocimiento: si se conocieran todas las fuerzas que actúan en un momento dado, se podría predecir exactamente el resultado. El modelo estadístico divide estas causas en dos grandes grupos: el primero lo forman un limitado número de fuerzas con entidad suficiente como para que su efecto pueda ser conocido y modelado de forma determinista; el segundo grupo, en cambio, es ilimitado y con influencia reducida: son tantas, y con efectos tan pequeños, que la única manera de modelar el resultado de su influencia es mediante la teoría de la combinatoria y de la probabilidad.

En este capítulo se introducen, con la ayuda de ejemplos, conceptos necesarios para interpretar los resultados de una prueba diagnóstica o la capacidad predictiva de un indicador.

A su vez, mediante el uso de la probabilidad condicionada se introduce al alumno en el concepto de riesgo y en las medidas de comparación de riesgos. Para aquellas situaciones en las que varía el tiempo de seguimiento, se introducen las tasas.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Interpretará el riesgo como una probabilidad.
- Interpretará la probabilidad de A condicionada a B como la probabilidad de A en el conjunto de casos que cumplen B.
- Distinguirá entre $P(A|B)$ y $P(B|A)$.
- Distinguirá entre sensibilidad, especificidad y valores predictivos.
- Definirá sensibilidad, especificidad y valores predictivos en términos de probabilidad condicionada.
- Para interpretar los resultados de una prueba diagnóstica, se preguntará por la prevalencia de la enfermedad en estudio.
- Sabrá que el teorema de Bayes también puede usarse con *odds*.
- Sabrá calcular un riesgo.
- Cuando los tiempos de seguimiento o el nivel de exposición varíen, calculará la tasa.
- Distinguirá entre riesgo y tasa.
- Distinguirá entre riesgo y *odd*.
- Sabrá calcular e interpretar el riesgo relativo, el atribuible y el *odds ratio*.
- En un estudio que fije los totales de enfermos y controles, usará el *odds ratio* y no el riesgo relativo ni el atribuible.

Probabilidad

Introducción a la probabilidad

Definición



La **probabilidad**, desde el punto de vista subjetivo, es el grado de certeza que se posee de un suceso.

Comentario



Se han propuesto varias definiciones de probabilidad. La primera definía la probabilidad como el «cociente entre casos favorables y casos posibles». Dada su circularidad (exige equiprobabilidad) enseguida se propuso sustituirla por «la frecuencia de aparición de un suceso». En el fondo, ambas definiciones son métodos para asignar valores a la probabilidad, es decir, para estimarla y conocer sus valores, pero no definiciones formales de lo que es en sí misma la probabilidad. En el siglo pasado se propusieron, casi al mismo tiempo, dos definiciones muy interesantes. Una de ellas, que era abstracta y axiomática, permitió un desarrollo formidable de toda la teoría de probabilidad. La otra definición, la subjetiva, permite expresar en términos de probabilidad tanto el grado de creencia en una teoría o en una afirmación científica, como la simple expectativa de la ocurrencia de un suceso. Su influencia en los conceptos de ciencia y de conocimiento aún está por desarrollarse plenamente.

Una representación gráfica del riesgo o probabilidad del suceso A, $P(A)$, es la siguiente (fig. 3-1), donde su valor concreto es directamente el cociente entre el área sombreada (A) y el área total (Ω).

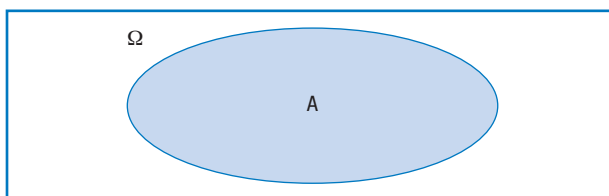


Figura 3-1 Probabilidad del suceso A.

Ejemplo 3.1



En 1978 se registraron 66 muertes por leucemia en una población de 890.575 individuos.

$P = \text{muertes por leucemia} / \text{población total} = 66 / 890.575 = 0'0000741$

Si se estima la probabilidad de morir por leucemia mediante el estimador proporción, se obtiene un riesgo de 74 muertes por millón de habitantes.

Recuerde



El riesgo es una probabilidad, por lo que no tiene unidades de medida y en el denominador están incluidos todos los casos del numerador.

Probabilidad condicionada

Existen infinitud de factores que pueden modificar esta probabilidad («las fuerzas con entidad suficiente»). Para considerarlas es preciso realizar la siguiente definición.

Definición



Probabilidad del suceso A condicionado al suceso B es la probabilidad de aparición del suceso A en los casos que cumplen también la característica B. Se representa por $P(A|B)$.

En el gráfico siguiente (fig. 3-2) su cálculo se obtiene mediante el cociente del área común a ambos sucesos ($A \cap B$) y el área del suceso B. Así, la probabilidad de A condicionada a B, $P(A|B)$, se puede entender como una reducción de la población en estudio: en lugar de considerar todos los casos de Ω , se reduce a los casos que cumplen la característica B.

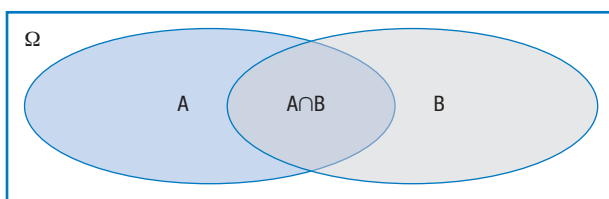


Figura 3-2 Probabilidad del suceso A condicionado al suceso B.

Recuerde



*$P(A|B)$ representa la probabilidad del suceso A «dentro» del total de casos que cumplen B.
 $P(A \cap B)$ representa la probabilidad de que ocurran simultáneamente los sucesos A y B.*

Definición



$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{Probabilidad de ser, a la vez, A y B}}{\text{Probabilidad de ser B}}$$

Ejercicio 3.1

¿Qué sería en el gráfico $P(B|A)$? ¿Cuál sería su definición?

Aplicación de la probabilidad al diagnóstico**Lectura**

Bossuyt et al. (19) ponen como ejemplo de una buena definición de objetivos en un estudio de capacidad diagnóstica: «Objetivo: determinar la sensibilidad y la especificidad de la colonografía tomográfica computarizada aplicada a la detección de los pólipos y el cáncer colorrectales tomando la colonoscopia como estándar de referencia». Y añaden: «Los estudios de precisión diagnóstica tienen una estructura básica común. Se evalúan una o más pruebas con el objetivo de detectar o predecir una cierta condición. La condición objetivo puede ser una enfermedad específica, el estadio de una enfermedad, un nivel de salud o cualquier otra condición del paciente. [...] En este contexto, “prueba” se refiere a cualquier método que permita obtener información adicional sobre el nivel de salud del paciente. Entre estas pruebas están los análisis de laboratorio, las técnicas de imagen, las determinaciones funcionales, el estudio anatomopatológico, la historia clínica y la exploración física. En un estudio de precisión diagnóstica, la prueba evaluada (denominada en este caso prueba índice) se aplica a una serie de individuos. Los resultados obtenidos con la prueba índice se comparan con los correspondientes al estándar de referencia, obtenidos en los mismos individuos. En este contexto, el estándar de referencia es el mejor método existente para establecer la presencia o ausencia de la condición objetivo y puede constar de una única prueba o una combinación de métodos y técnicas, incluyendo el seguimiento clínico de los individuos evaluados. El término de precisión se refiere al grado de concordancia entre los resultados obtenidos con la prueba índice y los obtenidos con el estándar de referencia. La precisión diagnóstica se puede expresar de distintas maneras, como el binomio sensibilidad-especificidad, la razón de verosimilitud, el cociente de posibilidades diagnósticas y el área bajo la curva ROC [receiver-operating characteristic]».

Por simplicidad, supóngase que se debe realizar el proceso diagnóstico de una sola enfermedad con dos únicos posibles estados, enfermo (E) y sano (S), y que se dispone de un único indicador con dos posibles valores, positivo (+) y negativo (–).

Recuerde

La primera dificultad consiste en definir qué variable mide perfectamente, sin error, el estado enfermo / sano; y qué variable representa el resultado del test positivo / negativo.

Definición



Sensibilidad (*Sens*): tendencia o propensión de los enfermos a dar positivo (en esta prueba).

Especificidad (*Esp*): tendencia o propensión de los sanos a dar negativo.

Valor predictivo positivo (*VP+*): confianza o credibilidad de un resultado positivo (de esta prueba).

Valor predictivo negativo (*VP-*): confianza o credibilidad de un resultado negativo.

Ejemplo 3.2



Una prueba sería extremadamente sensible si, aplicada a un conjunto de enfermos, casi el 100% dan positivo. Y un resultado tendrá un elevado valor predictivo positivo si casi el 100% de los que dan positivo están realmente enfermos. ¡Parece lo mismo, pero no lo es! La probabilidad condicionada, como se verá, ayuda a distinguirlos.

Ejemplo 3.3



Gisbert et al. (20) obtienen los siguientes resultados: «se incluyó a 117 pacientes con enfermedad de Crohn (EC), 72 con colitis ulcerosa (CU) y 2 con colitis indeterminada. Un paciente con EC (0,9%) y 6 con CU (8,3%) presentaron positividad en la determinación de PANCA. La sensibilidad, la especificidad, los valores predictivos positivo y negativo de los PANCA para el diagnóstico de CU (en comparación con la EC) fueron del 8, el 99, el 86 y el 64% respectivamente, lo que indica la necesidad de estandarizar la metodología. Conclusiones: la sensibilidad para el diagnóstico de CU en los pacientes con enfermedad inflamatoria intestinal es muy baja, si bien la especificidad es muy elevada».

Utilizando la probabilidad condicionada, se pueden definir estos conceptos con mucha mayor formalidad, aclarando sus diferencias y similitudes. De hecho, aparecen cuatro probabilidades condicionadas muy interesantes: $P(E|+)$ o probabilidad de enfermo si ha dado positivo, $P(+|E)$ o probabilidad de positivo en los enfermos, $P(-|S)$ o probabilidad de negativo en los sanos y $P(S|-)$ o probabilidad de sano en los negativos. Y también pueden ser interpretados usando la proporción muestral en lugar de la probabilidad poblacional: $P(S|+)$ o proporción de positivos que están sanos, $P(+|E)$ o proporción de enfermos que dan positivo, $P(-|S)$ o proporción de sanos que dan negativo y $P(S|-)$ o proporción de negativos que están sanos.

Ejercicio 3.2

Haga corresponder cada uno de estos términos con los conceptos anteriores.

Ejercicio 3.3

Intente explicar con sus propias palabras qué significan la especificidad y el valor predictivo negativo.

En la sensibilidad y en la especificidad, el condicionante, o punto de salida, es la realidad (enfermo o sano); mientras que el condicionado, o punto de llegada, es el indicador (positivo o negativo). Así, ambas van de la realidad al resultado: siguen el planteamiento «racionalista» de ir de la causa a la consecuencia. Los valores predictivos, en cambio, van al revés: condicionan o parten del indicador y llegan al estado real del paciente. Por ello contestan la pregunta «empirista» del diagnóstico: dado que este paciente ha presentado estos síntomas, signos e indicadores, ¿está enfermo? O mejor: ¿qué probabilidades tiene de estar realmente enfermo sabiendo que ha dado positivo, $P(E|+)$? ¿O de estar sano si ha dado negativo, $P(S|-)$?

Ejemplo 3.4

La sensibilidad, la especificidad y los valores predictivos de la tabla 3-1, que se muestra a continuación son:

	+	-	TOTAL	
Enfermo	94	38	132	$Sens = P(+ E) = 94/132 \approx 0,712 = 71,2\%$
Sano	215	653	868	$Esp = P(- S) = 653/868 \approx 0,752 = 75,2\%$
TOTAL	309	691	1.000	$VP+ = P(E +) = 93/309 \approx 0,304 = 30,4\%$
				$VP- = P(S -) = 653/691 \approx 0,945 = 94,5\%$

Tabla 3-1 Probabilidades diagnósticas en una muestra con un 13,2% de enfermos

A pesar de que los valores de sensibilidad y especificidad son muy similares, los valores predictivos se alejan considerablemente entre sí: la probabilidad de que un paciente que dé negativo esté sano ($VP-$) es alta, pero la probabilidad de que un paciente que dé positivo esté enfermo ($VP+$) es baja. La razón es muy simple: ya que hay más sanos (86,8%) que enfermos (13,2%), la segunda fila pesa más que la primera.

Ejemplo 3.5



Tal y como puede observarse en la tabla 3-2, si cambiamos la $P(E)$ de 0,132 a 0,75, pero mantenemos una sensibilidad y una especificidad iguales a las anteriores, los valores predictivos cambian.

	+	-	TOTAL
Enfermo	534	216	750
Sano	62	188	250
TOTAL	596	404	1.000

$Sens = P(+/E) = 537/750 \approx 0,712 = 71,2\%$

$Esp = P(-/S) = 188/250 \approx 0,752 = 75,2\%$

$VP+ = P(E/+) = 534/596 \approx 0,896 = 89,6\%$

$VP- = P(S/-) = 188/404 \approx 0,465 = 46,5\%$

Tabla 3-2 Probabilidades diagnósticas en una muestra con un 75% de enfermos

No debe olvidarse cuál es la proporción de enfermos en la población origen de los datos. Si es muy baja, la proporción de enfermos seguirá siendo baja después del resultado de la prueba, lo que llevará a un valor predictivo positivo (proporción de enfermos en los positivos) muy bajo y a un valor predictivo negativo (proporción de sanos en los negativos) muy alto. Un observador poco atento se podría olvidar de las condiciones de salida y esperar que, si un 70% de los enfermos dan positivo, pues aproximadamente un 70% de los positivos estarán enfermos.

La sensibilidad y la especificidad son el resultado de un mecanismo fisiológico o patológico y, por tanto, suelen ser extrapolables de una población de pacientes a otra. En cambio, los valores predictivos también dependen de la frecuencia de la enfermedad en la población, por lo que variarán mucho de unas condiciones a otras.

Recuerde



La sensibilidad y la especificidad pueden ser universales, pero los valores predictivos dependen de la frecuencia de la enfermedad en la población en estudio.

Ejemplo 3.6



Suponga dos poblaciones diferentes: 1) la formada por los toxicómanos y homosexuales de una cierta prisión, y 2) la formada por los médicos de un cierto hospital. Suponga un resultado positivo a la misma prueba del sida de dos miembros de cada una de estas dos comunidades. ¿Tienen es-

Ejemplo 3.6 (Cont.)

tos dos individuos las mismas probabilidades de padecer sida? La respuesta es no: delante de un resultado positivo a una prueba del sida, el valor predictivo positivo o probabilidad de estar realmente enfermo habiendo dado positivo es mucho mayor en el primer caso. Más adelante se verá una exposición formal de este hecho aparentemente paradójico.

Comentario

Todos estos fenómenos paradójicos tienen menor repercusión cuanto menos aleatorio es el fenómeno en estudio. Si la sensibilidad y la especificidad son ambas del 100%, no hay duda de que un positivo está enfermo y un negativo está sano.

Ejercicio 3.4

Construya la tabla 2×2 del ejemplo anterior con los números correspondientes a esta nota (sensibilidad = especificidad = 100%) y compruebe que la afirmación realizada en la nota es cierta (suponga que de 1000 casos estudiados, estaban enfermos 132 y 868, sanos). ¿Se cumple que ambos valores predictivos son del 100%?

Ejercicio 3.5

Mirando a la tabla construida en el ejercicio anterior, para que el valor predictivo positivo sea perfecto ($VP+ = 100\%$), ¿qué se necesita que sea del 100%, la sensibilidad o la especificidad?

Ejercicio 3.6

En clínica se dice que un signo es patognomónico cuando su presencia asegura definitivamente la enfermedad que ayuda a diagnosticar. ¿Alguna de las probabilidades diagnósticas es del 100%?

Ejercicio 3.7

Intente explicar con sus propias palabras lo que miden la sensibilidad, la especificidad y ambos valores predictivos.

Ejercicio 3.8

Repita los cálculos, suponiendo que se habían obtenido dos muestras de 1.000 casos, una para enfermos y otra para sanos. Tabla 3-3.

	+	-	TOTAL
Enfermo	712	288	1.000
Sano	248	752	1.000
TOTAL	960	1.040	

Tabla 3-3 Probabilidades diagnósticas en dos muestras de 1.000 casos

En el ejercicio anterior ambas filas pesan lo mismo, y los valores predictivos se han parecido más a la sensibilidad y especificidad. Pero este peso similar de ambas filas se ha fijado artificialmente, por lo que estos valores predictivos sólo se corresponderían con una realidad en la que ambas filas pesaran lo mismo. Cuando el número de enfermos y de sanos está fijado por muestreo, el cálculo de los VP no puede hacerse con el procedimiento anterior y debe usarse el que se explica en el siguiente apartado.

Recuerde



La mayor parte de diseños para calcular la capacidad de una prueba diagnóstica se basan en una muestra de enfermos obtenida de un determinado servicio y una muestra de sanos obtenida por otro medio. Dado que el número de casos en cada fila lo decide el investigador, no se pueden calcular directamente los valores predictivos.

Teorema de Bayes

El teorema de Bayes permite resolver el problema anterior, ya que incorpora la información sobre la proporción de enfermos que hay en la población en estudio. Esquemáticamente, consiste en combinar los valores de sensibilidad y especificidad con esta proporción de enfermos para obtener los valores predictivos (fig. 3-3).

Veamos en primer lugar la fórmula del teorema de Bayes.

Definición



Sean A y B dos eventos cualesquiera (por ejemplo A=ser rubio y B=ser nórdico)

$$\text{Teorema de Bayes: } P(B|A) = \frac{P(A|B) P(B)}{P(A)}$$

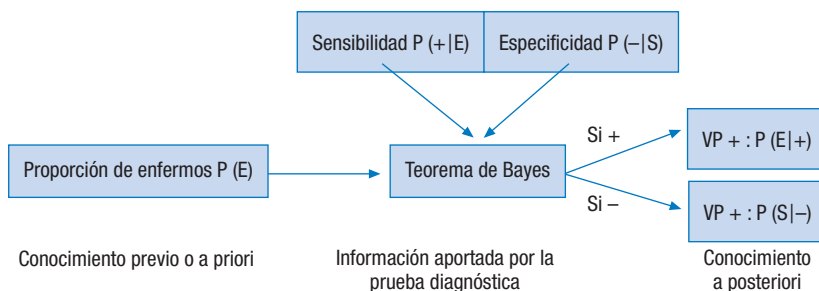


Figura 3-3 Esquema del teorema de Bayes.

Nótese que el teorema de Bayes permite invertir los términos de condicionante y condicionado: a partir de $P(A|B)$ se obtiene $P(B|A)$, y por tanto, a partir de $P(+|E)$ se llega a $P(E|+)$.

Nota técnica



Demostración a partir de la definición de probabilidad condicionada

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$$

$$\text{Y entonces } P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Comentario



El monje Bayes quería demostrar la existencia de Dios (D). Partía de que el ser humano existe [$P(H)=1$] y de que Dios, por su propia definición, crearía la humanidad [$P(H|D) = 1$]. Su pregunta era: ¿cuál es la probabilidad de que exista Dios, sabiendo que el ser humano existe, [$P(D|H)$]? Y Bayes, para contestar a esta pregunta, desarrolló el teorema que lleva su nombre. La respuesta que alcanzó fue:

$$P(R|H) = \frac{P(H|D) \cdot P(D)}{P(H)} = \frac{1 \cdot P(0)}{1}$$

Bayes observó que todo dependía de sus probabilidades iniciales de creer en Dios. Si él, a priori ya creía en Dios, $P(D)=1$ entonces, habiendo observado que existía el hombre, también $P(D|H)=1$. En cambio, si a priori no lo hacía: $P(D)=0$, tampoco a posteriori; tras observar al hombre: $P(D|H)=0$. Por lo que su teorema no servía para demostrar la existencia de Dios. Thomas Bayes, al no conseguir sus objetivos, olvidó sus resultados, que sólo fueron publicados póstumamente por el prior que le sucedió.

Nota técnica



La fórmula del teorema de Bayes se puede expresar también mediante la descomposición del denominador A en los dos sucesos $A \cap B$ y $A \cap B^c$ que lo componen (B^c representa el contrario o negación de B; si, por ejemplo, B representa enfermo, B^c representará sano):

$$\begin{aligned} P(B|A) &= \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A \cap B) + P(A \cap B^c)} = \\ &= \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)} \end{aligned}$$

Recuerde



Una expresión alternativa del teorema de Bayes es:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}$$

Ejemplo 3.7



Las sensibilidades y especificidades obtenidas en el ejercicio 3.8 se quieren aplicar en un entorno en el que hay una $P(E) = 0,132$. El valor predictivo positivo será:

$$\begin{aligned} P(E|+) &= \frac{P(+|E) \cdot P(E)}{P(+|E) \cdot P(E) + P(+|S) \cdot P(S)} \\ &= \frac{0,712 \cdot 0,132}{0,712 \cdot 0,132 + 0,248 \cdot 0,868} \approx 0,304 \end{aligned}$$

Ejercicio 3.9



Compruebe que sabe calcular el valor predictivo negativo.

El teorema de Bayes permite, a partir de la proporción de enfermos, $P(E)$, de una población y de la sensibilidad y especificidad de una prueba, obtener los valores predictivos que se pueden aplicar a un paciente determinado.

Comentario



En ocasiones, se habla de **falsos positivos** (FP) y de **falsos negativos** (FN), que representan los complementarios o contrarios de la especificidad y sensibilidad, respectivamente. Es decir, si la sensibilidad es del 90%, quiere decir que un 10% de los enfermos dan un resultado negativo que es falso.

Ejemplo 3.8

Una prueba diagnóstica para la diabetes tiene FP de 4% y FN del 5%. Si la prevalencia de la diabetes en la población donde se usa es del 7%, ¿cuál es la probabilidad de que sea diabético un individuo en el que la prueba dé positivo? ¿Y de que no lo sea uno en el que dé negativo?

$$p(+|S) = 0,04 \rightarrow p(-|S) = 1 - 0,04 = 0,96 = \text{Esp}$$

$$p(-|E) = 0,05 \rightarrow p(+|E) = 1 - 0,05 = 0,95 = \text{Sens}$$

$$p(E) = 0,07 \rightarrow p(S) = 1 - 0,07 = 0,93$$

$$P(E|+) = \frac{P(+|E) \cdot P(E)}{P(+|E) \cdot P(E) + P(+|S) \cdot P(S)} =$$

$$= \frac{0,95 \cdot 0,07}{0,95 \cdot 0,07 + 0,04 \cdot 0,93} = 0,641 = VP^+$$

$$P(S|-) = \frac{P(-|S) \cdot P(S)}{P(-|S) \cdot P(S) + P(-|E) \cdot P(E)} =$$

$$= \frac{0,96 \cdot 0,93}{0,96 \cdot 0,93 + 0,05 \cdot 0,07} = 0,996 = VP^-$$

El teorema de Bayes permite ir actualizando la información de que se dispone: su fórmula «mezcla» la información previa, disponible a priori $P(E)$, con los nuevos resultados (+ o -).

Recuerde

$$\text{Información a priori} + \text{nueva información} = \text{Información a posteriori}$$

Riesgos y tasas**Riesgo**

En un sentido amplio, el riesgo es la probabilidad de que algo ocurra. En investigación clínica, es el resultado desfavorable de una actividad, intervención o exposición, especialmente referido a la probabilidad de que aparezca un fenómeno adverso concreto.

Lectura

En el caso de estudios sin beneficio terapéutico, se acepta como lícito el hecho de que los voluntarios se encuentren sometidos a riesgo mínimo o insignificante, según la FDA. Se acepta como tal la probabilidad de entre 1 y 100 por mil de sufrir una complicación menor y de entre 10 y 1.000 por millón de sufrir una complicación grave (21, p. 56).

Ejemplo 3.9



Suponga que el diagnóstico de una enfermedad (Y) y la existencia de un Factor de Riesgo (X) sólo pueden tomar dos valores (es decir, son «dicotomías») que son: presente (+) y ausente (-). Así, por ejemplo, Y+ representará que tiene la enfermedad y X- que no está expuesto al factor de riesgo. El riesgo absoluto de sufrir la enfermedad en estudio sería $P(Y+)$. El riesgo entre los expuestos sería: $P(Y+ | X+)$. El riesgo entre los no expuestos sería $P(Y+ | X-)$. Es habitual presentar simultáneamente estas dos dicotomías en forma de una tabla 2×2 , en las que los marginales representan los totales de cada variable. Así, en la tabla 3-4 puede leerse que de los 1.000 casos estudiados, 15 presentaban la enfermedad, de los que 7 estaban expuestos y 8 no.

Ahora se pueden calcular los riesgos habituales con las fórmulas de probabilidad condicionada:

Riesgo: $P(Y+) = 15 / 1.000 = 0,015$

Riesgo en los Expuestos: $P(Y+|X+) = 7 / 132 \approx 0,053$

Riesgo en los No-Expuestos: $P(Y+|X-) = 8 / 868 \approx 0,009$

	Y+	Y-	TOTAL
X+	7	125	132
X-	8	860	868
TOTAL	15	985	1.000

Tabla 3-4 Presencia de la enfermedad Y y el factor de riesgo X en 1.000 casos

Ejercicio 3.10



Fumaz et al. (22) comparan dos grupos de pacientes tratados, uno con efavirenz (EFV, $n = 51$) y otro con inhibidores de la proteasa (PI, $n = 49$), habiendo observado respectivamente 36 y 9 acontecimientos adversos relacionados con el sistema nervioso central. Elabore la tabla 2×2 y calcule los riesgos respectivos.

Tasa

En muchas ocasiones, se observa a los pacientes bajo diferentes circunstancias. La más habitual es que diversos casos hayan sido estudiados durante un diferente tiempo de seguimiento. Por ello, no es lo mismo que el evento de interés lo haya experimentado un caso seguido durante 1 mes que uno seguido durante 2 años. Conviene considerar esta nueva información a la hora de calcular los riesgos. Para ello, se debe incluir en el denominador el tiempo total de seguimiento de cada caso, con lo que el

riesgo así calculado es el cociente entre un número de casos y una suma de tiempos de seguimiento, por lo que ya no se trata de una probabilidad, que eran casos (posibles) divididos entre casos (totales). Por ello, se habla de tasas. En los temas específicos de supervivencia se estudian con más detalle.

Definición



Tasa es la medida de la frecuencia de un evento expresada de forma relativa al tiempo.

Ejemplo 3.10



Regidor et al. (23). En 1998 se produjeron en España 360.511 defunciones, lo que supone una tasa de mortalidad de 915,7 por 100.000 habitantes en un año de seguimiento. En la tabla 3-5 aparecen ordenadas jerárquicamente, en virtud del número de fallecimientos, las 12 causas de muerte estudiadas. Esas causas ocasionaron el 78% de todos los fallecimientos ocurridos en España ese año. Los tumores malignos fueron la principal causa de muerte, con una tasa de mortalidad de 227,7 por 100.000 habitantes-año, lo que supone el 24,9% de todas las defunciones; las enfermedades del corazón figuraban en segundo lugar, con una tasa de 207,7 por 100.000 y el 22,7% de los fallecimientos, y las enfermedades cerebrovasculares en el tercero, con una tasa de 96,8 por 100.000 y el 10,6% de las defunciones.

Odd

Los países de tradición anglosajona también usan una forma alternativa a la probabilidad para expresar los resultados inciertos. Si la probabilidad expresa los casos a favor divididos entre todos los casos posibles, las *odds* hablan de los casos a favor divididos entre los casos en contra.

Ejemplo 3.11



Así, mientras nosotros diríamos que cierto caballo tiene 7 números sobre 8 de ganar una carrera, los anglosajones podrían decir también que los números de este caballo están 7 a 1, indicando que tienen 7 a favor y 1 en contra.

Comentario



En los ambientes de apuestas (pelota vasca, carreras de galgos, etc.) se emplea un término equivalente a las odds. Se dice, por ejemplo, que las apuestas por cierto pelotari están 7 momios a 1. ¿Alguien se atreve a traducir odd por momio?

Causas de muerte (CIEB.* revisión)	Defunciones	Tasas brutas	Mortalidad proporcional (%)	Porcentaje de cambio en la tasa ajustada de mortalidad	
				1995-1998	1980-1988
Todas las causas	360.511	915,7	100,0	-3,2*	-20,2*
Cáncer (140-208)	89.665	227,7	24,9	-2,4*	7,7*
Enfermedades del corazón (390-398, 410-429)	81.768	207,7	22,7	-1,8*	-26,4*
Enfermedad cerebrovascular (430-438)	38.121	96,8	10,6	-12,0*	-52,8*
Enfermedad pulmonar obstructiva crónica (490-796)	17.768	45,1	4,9	1,1 NS	30,8*
Accidentes no intencionales (E800-E949)	13.122	33,3	3,6	0,1 NS	-21,6*
Diabetes mellitus (250)	9.533	24,2	2,6	-3,4***	-15,1*
Neumonía e influenza (480-487)	8.491	21,6	2,4	3,9***	-52,6*
Cirrosis y otras enfermedades crónicas del hígado (571)	6.246	15,9	1,7	-14,3*	-45,3*
Nefritis, síndrome nefrótico y nefrosis (580-589)	5.566	14,1	1,5	-1,6 NS	-16,1*
Aterosclerosis (440)	4.717	12,0	1,3	-26,5*	-81,2*
Enfermedad de Alzheimer (331,0)	3.551	9,0	1,0	42,9*	2.111,6*
Suicidio (E950-E959)	3.261	8,3	0,9	0,1 NS	53,8*

*p < 0,001; **p < 0,01; ***p < 0,05; NS: no significativo.

Tabla 3-5 Principales causas de muerte en España en 1998. Número de defunciones, tasas de mortalidad por 100.000 habitantes, mortalidad proporcional y porcentajes de cambio en la mortalidad ajustada por edad de 1995 a 1998 y de 1980 a 1998

Definición



La *odd* de A es la probabilidad de que se presente el suceso A dividida por la probabilidad de que no se presente A.

$$Odd(A) = \frac{P(A)}{P(\text{no } A)}$$

Ejercicio 3.11



¿Cuánto valen la probabilidad y la *odd* de sacar un «3» en el lanzamiento de un dado?

Ejemplo 3.12



Veamos cómo se calculan las *odds* en los mismos datos del ejemplo 3.9 (tabla 3-4):

Odd expuestos: $O(Y+ | X+) = 7 / 125 \approx 1 / 18 \approx 0,056$

Odd no expuestos: $O(Y+ | X-) = 8 / 860 \approx 1 / 107 \approx 0,0093$

En los expuestos, la enfermedad aparece en 1 caso por cada 18 en los que no aparece. En cambio, en los no expuestos, la enfermedad aparece en 1 caso por cada 107 en los que no.

Nota técnica

Si la probabilidad de enfermedad es muy pequeña, la probabilidad de estar sano será muy próxima a 1, por lo que la *odd* tendrá un valor muy similar a la probabilidad:

$$Odd(\text{enfermo}) = \frac{P(\text{enfermo})}{P(\text{sano})} \approx \frac{P(\text{enfermo})}{1} = P(\text{enfermo})$$

Ejercicio 3.12

El dolor lumbar o la gripe son enfermedades comunes en el sentido de que a lo largo de la vida es fácil padecerlas al menos en una ocasión. Pongamos que sus probabilidades respectivas son 0,5 y 0,8. Calcule sus odds. La esclerosis múltiple, en cambio, es muy poco frecuente. Pongamos que la probabilidad de padecerla a lo largo de la vida es de 0,001 (uno por mil). Calcule la *odd*.

Recuerde

En el caso de enfermedades «raras», el riesgo y la odd dan resultados muy similares.

Ejemplo 3.12 (Cont.)

La *odd* en los expuestos vale 0,056 [$O(Y+ | X+) = 7/125 \approx 0,056$], valor muy similar al 0,053 anterior del riesgo en los expuestos.

En los no expuestos la similitud es aún mayor, ya que la *odd* vale 0,0093 por 0,0092 el riesgo.

Un aspecto importante a tener en cuenta es la forma en la que se han recogido los datos, ya que si, se ha *forzado* la recogida de datos para que tenga mayor representación cierto tipo de casos, algunos de los cálculos pueden no ser correctos.

Comentario

De la misma forma que los valores predictivos no podían calcularse si el diseño implicaba dos muestras, una de enfermos y una de sanos, el siguiente ejercicio evidencia que tampoco pueden calcularse los riesgos.

Ejercicio 3.13



En una cierta población, se ha recogido información en 1.000 casos sobre dos variables, la exposición al riesgo del tabaco (fumador: X+ y no fumador: X-) y su evolución posterior (bronquitis: Y+ y no bronquitis: Y-). A partir de los resultados de la tabla 3-6A, calcule los riesgos y las odds.

Repita los cálculos en la tabla 3-6B, en la que los datos se han obtenido de dos muestras de 1000 casos, una de fumadores y otra de no fumadores.

Repita una vez más, tabla 3-6C, suponiendo ahora que las dos muestras de 1.000 casos corresponden una a bronquíticos y otra a no bronquíticos.

	Y+	Y-	TOTAL
X+	94	38	132
X-	215	653	868
TOTAL	309	691	1.000

Tabla 3-6 A Datos sobre una única muestra de 1.000 casos

	Y+	Y-	TOTAL
X+	712	288	1.000
X-	248	752	1.000
TOTAL	960	1.040	

Tabla 3-6 B Datos sobre dos muestras de 1.000 casos, una de fumadores (X+) y otra de no fumadores (X-)

	Y+	Y-	TOTAL
X+	304	55	359
X-	696	945	1.641
TOTAL	1.000	1.000	

Tabla 3-6 C Datos sobre dos muestras de 1.000 casos, una de pacientes con bronquitis (Y+) y otra sin (Y-)

Recuerde

Si se fija por diseño el número total de casos con la enfermedad y el número de casos sin la enfermedad, ya no pueden calcularse ni los riesgos ni las odds de desarrollar la enfermedad (ni en los expuestos ni en los no expuestos).

Odds diagnóstica

Una de las ventajas de trabajar con odds en lugar de con probabilidades es que el teorema de Bayes se simplifica mucho. Como explica la nota técnica siguiente, conocer las *odds* de enfermo a sano tras la obtención de un valor positivo del indicador diagnóstico consiste en multiplicar la **razón de verosimilitud** (*likelihood ratio*), de un resultado positivo (sensibilidad dividida entre el complementario de la especificidad) por la *odd* previa de enfermo a sano en un entorno determinado.

Nota técnica

Utilizando probabilidades condicionadas se obtuvo que:

$$P(E|+) = P(+|E) P(E)/P(+) \quad \text{y} \quad P(S|+) = P(+|S) P(S)/P(+)$$

Recordando la definición de la *odd* $(E) = P(E)/P(S)$, y aplicándola a las probabilidades condicionadas anteriores, se pueden obtener las *odds* a posteriori entre enfermo (E) y sano (S), una vez conocido el resultado positivo de la prueba:

$$\underbrace{\frac{P(E|+)}{P(S|+)}}_{\substack{\text{Odds} \\ \text{a} \\ \text{posteriori}}} = \frac{P(+|E) P(E)/P(+)}{P(+|S) P(S)/P(+)} = \underbrace{\frac{P(+|E)}{P(+|S)}}_{\substack{\text{Razón de} \\ \text{verosimi-} \\ \text{litud (RV)}}} \cdot \underbrace{\frac{P(E)}{P(S)}}_{\substack{\text{Odds} \\ \text{a} \\ \text{priori}}}$$

Ejemplo 3.4 (Cont.)

La razón de verosimilitud (RV) de un resultado positivo vale

$$\frac{P(+|E)}{P(+|S)} = \frac{0,712}{1-0,752} \approx 2,87$$

Si se sabe que en cierto servicio sanitario hay 1 enfermo por cada 3 sanos (*odds a priori*), una vez haya dado positivo el indicador diagnóstico, habrá aproximadamente 1 enfermo por cada 1 sano (*odds a posteriori* = *RV* × *odds a priori* = $2,87 \times 1/3 = 0,96 \rightarrow 0,96$ enfermos por cada 1 sano ≈ 1 enfermo por cada sano).

Ejercicio 3.14



Suponga que el resultado de la prueba en el ejemplo anterior ha sido negativo, ¿Cuánto vale la RV de enfermo para un resultado negativo? ¿Cuánto valen las *odds* a posteriori de enfermo para un resultado negativo?

Nota técnica



Si se desea trabajar de forma aditiva, se puede recurrir a tomar logaritmos

$$\underbrace{\log(Odds_{posteriori})}_{\text{Información a posteriori}} = \underbrace{\log(RV)}_{\text{Función SOPORTE}} + \underbrace{\log(Odds_{priori})}_{\text{Información a priori}}$$

Es decir, los *logodds* a priori más la información aportada por el resultado empírico («función soporte»), proporcionan los *logodds* a posteriori.

Ya se ha dicho que las *odds* («casos a favor» / «casos en contra») son una forma alternativa de presentar la probabilidad («casos a favor» / «casos totales») que suele emplearse en la cultura anglosajona en los entornos de apuestas. Ahora se ha visto que conduce a cálculos más simples y directos, lo que representa su principal ventaja. En el libro de Guyat et al. (24) pueden verse numerosos ejemplos de su aplicación en la práctica clínica.

Incidencia y prevalencia

A continuación se definen dos medidas básicas de frecuencia de morbilidad.

Definición



La incidencia se refiere al número de casos nuevos de una enfermedad en una población durante un período de tiempo determinado.

Ejemplo 3.13



Cohn et al. (25). La incidencia del criterio de valoración fue un 13,2% más baja con Valsartán que con placebo (riesgo relativo, 0,87).

Definición

La prevalencia define el número de individuos que presenta una determinada característica o enfermedad en una población y en un momento de tiempo determinado.

Ejemplo 3.14

Martín et al. (26). El asma es un grave problema de salud en los países industrializados, donde supone una de las enfermedades crónicas más frecuentes, sobre todo en la infancia. En Europa, la prevalencia varía ampliamente entre los distintos países, con cifras que oscilan entre un 8% en el Reino Unido y un 2% en Grecia. También existen amplias diferencias entre distintas zonas del mismo país, encontrándose en España cifras entre el 5 y el 1%.

Ejemplo 3.15

Ricart et al. (27). Recientemente, la American Diabetes Association (ADA) concluyó que no se deben realizar estudios para detectar la diabetes mellitus gestacional (DMG) en el grupo de mujeres gestantes con bajo riesgo para desarrollar esta enfermedad. El objetivo de este trabajo es determinar en una población española la prevalencia de DMG en un grupo de gestantes con bajo riesgo [...]. Diseño y métodos: revisión de una cohorte de 2.262 gestaciones (2.085 caucásicas), controlada desde 1990 a 1998. Se estudia la prevalencia y las características de las gestantes de bajo riesgo. Resultados: la prevalencia de DMG fue del 15%. Doscientas setenta y cuatro mujeres (12,1%) se catalogaron de bajo riesgo de desarrollar una DMG, 13 de las cuales (4,7%) presentaban DMG, que contrasta con el 16,6% en el resto de las mujeres estudiadas.

Comentario

Prevención: cualquier intervención que reduzca el riesgo de que una enfermedad o trastorno afecte a un individuo, que interrumpa o detenga su progreso o evite la muerte. Podemos distinguir:

- Prevención **primaria:** cualquier intervención dirigida a individuos sanos sin la enfermedad cuya aparición se quiere evitar.
- Prevención **secundaria:** desde un punto de vista diagnóstico, intervenciones dirigidas a detectar precozmente una enfermedad.
- Prevención **terciaria:** intervención preventiva en pacientes que ya han padecido una enfermedad relacionada con la rehabilitación y la mejora de la calidad de vida.

Ejemplo 3.16



- De la prevención primaria: vacunaciones, modificación de factores de riesgo como el tabaco para evitar el cáncer o la utilización del ácido acetilsalicílico para prevenir el infarto de miocardio en sujetos sanos.
- De la prevención secundaria: la utilización de la mamografía para la detección del cáncer de mama o la prueba de Papanicolaou para el cáncer de cérvix.
- De la prevención terciaria: tratamiento fisioterapéutico y rehabilitador en pacientes que han padecido un accidente cerebrovascular.

Medidas de asociación en tablas 2×2

Una vez estimado el riesgo de desarrollar una enfermedad en un grupo determinado puede ser interesante la comparación de estos riesgos entre dos grupos de diferentes características.

Ejemplo 3.9 (Cont.)



Habíamos visto en el **ejemplo 3.9**, al hablar de probabilidad condicionada, que el riesgo o probabilidad en los expuestos era de un 5,3% [$P(Y+|X+) = 7 / 132 \approx 0,053$], mientras que en los no expuestos era del 0,9% [$P(Y+|X-) = 8 / 868 \approx 0,009$] (tabla 3-4).

Estos dos números se pueden comparar matemáticamente mediante la resta o la división. Así, el riesgo atribuible consiste en hacer la diferencia entre el riesgo en los expuestos y los no expuestos, mientras que el riesgo relativo es su cociente.

Lectura



CONSORT (9): «Para cada criterio de valoración principal y secundario. [...] En respuestas binarias, la medida del efecto podría ser el cociente de riesgo (riesgo relativo), el odds ratio o el riesgo atribuible; en lo relativo a los datos del tiempo de supervivencia, la medida del efecto podría ser la razón de tasas o la diferencia en la supervivencia mediana. En los datos continuos la medida del efecto suele ser la diferencia en las medias. [...]

En lo que se refiere a los datos binarios y supervivencia, también puede ser útil la expresión de los resultados en forma del número necesario de pacientes que se han de tratar para la obtención de efecto beneficioso (NNTB, number needed to treat for benefit) o de efecto adverso (NNTH, number needed to treat for harm).»

Riesgo atribuible

Definición



El *riesgo* de una enfermedad *atribuible* a una exposición es la diferencia entre el riesgo en los expuestos y el mismo riesgo en los no expuestos.

Ejemplo 3.16 (Cont.)



En el ejemplo anterior, la diferencia entre 0,053 y 0,009 es 0,044; es decir, expresado en porcentajes, un 4,4%.

El riesgo atribuible puede tomar el valor máximo de 1 (o 100%) en el caso de que todos los expuestos desarrollaran la enfermedad y, en cambio, no lo hiciera ninguno de los no expuestos. El valor 0 representa que el riesgo es el mismo en ambos grupos.

Nota



En teoría, sería posible observar valores negativos hasta un mínimo de -1 (-100%) en el caso de que el factor en estudio protegiera de la enfermedad en lugar de favorecer su aparición, aunque siempre es más fácil de interpretar si los resultados se expresan en sentido positivo, para lo que basta con intercambiar las definiciones de expuesto y no expuesto.

Nota técnica




El nombre de riesgo atribuible es muy ambicioso, ya que da a entender que este incremento del riesgo viene exclusivamente originado por la exposición al factor, para lo que es necesaria la existencia de una relación de causa-efecto. Establecer esta relación precisa toda una serie de consideraciones que se estudian más adelante y que van más allá de la simple medida de asociación entre dos variables. Mientras tanto, el concepto riesgo debe valorarse únicamente a nivel predictivo: los expuestos tienen ese mayor riesgo que los no expuestos.

Recuerde



El RA valora la diferencia de riesgos. Si la relación fuera causal, podría decirse que la exposición añade o suma el riesgo RA al riesgo de los no expuestos.


Riesgo relativo



Definición

El riesgo de una enfermedad en los expuestos *relativo* a los no expuestos es el cociente entre el riesgo en los expuestos y el mismo riesgo en los no expuestos.


Ejemplo 3.12 (Cont.)



En el **ejemplo 3.12** (cont. del 3.9), la razón entre 0,053 y 0,009 es prácticamente 6, lo que indica que los expuestos tienen un riesgo casi 6 veces superior a los no expuestos.


El riesgo relativo pretende evaluar cuánto se multiplica la probabilidad de desarrollar la enfermedad. Puede tomar el valor máximo de infinito (∞) en el caso de que ningún no expuesto desarrollara la enfermedad y, en cambio, sí que lo hiciera alguno de los expuestos. El valor 1 representa que el riesgo es el mismo en ambos grupos. Valores inferiores a 1 indicarían un factor protector.

Recuerde



El RR valora la razón de riesgos. Si la relación fuera causal, podría decirse que la exposición multiplica el riesgo RR veces.

Ejemplo 3.17



Banegas et al. (28). En la tabla 3-7 se muestra cómo van aumentando los RR de muerte cardiovascular. Nótese que el grupo «<120/80» se toma como referencia y, por tanto, $RR = 1$.

Presión arterial (mmHg)	Varones				Mujeres			
	%	Cardiopatía isquémica RR	Enfermedad cerebro-vascular (RR)	Todas las causas (RR)	%	Cardiopatía isquémica RR	Enfermedad cerebro-vascular (RR)	Todas las causas (RR)
<120/80	20,1	1	1	1	25,6	1	1	1
120-129/80-84	18,1	1,3	1,4	1,2	16	1,4	1,5	1,2
130-139/85-89	17,2	1,6	1,9	1,3	16,9	1,7	1,9	1,3
140-155/90-99	29,1	2,5	2,5	1,6	21,8	2,6	2,7	1,5
160-169/100-109	17,2	3,4	4,4	2,2	10,4	3,5	4,2	1,9
>180/110	3,3	4,5	6,3	3,4	3,3	4,5	6,2	3,4
TOTAL	100				100			

Tabla 3-7 Prevalencias de presión arterial en España y RR de muerte cardiovascular

Odds ratio

Definición



La *odds ratio* es el cociente entre las *odds* en los expuestos y las mismas *odds* en los no expuestos.

Ejemplo 3.12 (Cont.)



En el **ejemplo 3.12** (continuación del 3.9), la razón entre 0,056 y 0,009 es 6,2, lo que indica que la *odd* en los expuestos es 6,2 veces superior a la *odd* en los no expuestos.

Si la enfermedad es poco frecuente, la *odds ratio* es similar al riesgo relativo. En los ejemplos previos, 6,2 y 6, respectivamente. Algunas traducciones proponen llamarla riesgo relativo aproximado. Ambas medidas de asociación se interpretan de forma idéntica.

Recuerde



Riesgo relativo y odds ratio se interpretan igual.

Ejercicio 3.15



Calcule RR, RA y OR en las tablas del ejercicio 3.13. A pesar de que se ha dicho que no tenía sentido calcular los riesgos ni las *odds* en la tercera tabla, haga también en ella todos estos cálculos y observe qué sucede con el valor del *odds ratio*.

Una gran ventaja de la *odds ratio* sobre las medidas basadas en riesgos es que puede ser calculado en cualquier tabla 2×2 , independientemente de cuál haya sido el plan de muestreo.

Nota técnica



Ello es así porque, de la misma forma que se definió la OR como el cociente entre expuestos y no expuestos de las *odds* enfermo/sano $[(a/b)/(c/d)]$, también podría haberse definido como el cociente entre enfermos y sanos de las *odds* expuesto/no expuesto $[(a/c)/(b/d)]$. Dado que ambas definiciones son equivalentes, la *odds ratio* se puede utilizar en cualquier tabla de dos filas y dos columnas, independientemente del plan de muestreo: $(a/b)/(c/d) = (a/c)/(b/d) = ad/bc$

Nota técnica


Mientras que el RA descansa en un modelo aditivo (los riesgos se suman), en el RR subyace un modelo multiplicativo. Por ejemplo, por el hecho de fumar, el riesgo o probabilidad de tener bronquitis puede «sumarse» o «multiplicarse». El RA invita a pensar que ciertos casos siempre tendrán bronquitis (los que aparecen en los no expuestos), mientras que, por el hecho de fumar aparecen casos «nuevos». En cambio, el RR invita a pensar que cada caso tiene una cierta probabilidad de tener bronquitis (la proporción en los no expuestos), mientras que los fumadores tienen otra probabilidad multiplicada por RR.

Recuerde


La odds ratio tiene la gran ventaja de que se puede utilizar en cualquier estudio, independientemente del plan de muestreo. Esta propiedad no la tienen ni el riesgo atribuible ni el relativo.

Ejercicio 3.16


Imagine un diseño casos-controles, en los que se escoge una muestra de enfermos (casos) y una muestra de sanos (controles), por lo que se deja fija la variable enfermo / sano. ¿Cuáles de las medidas anteriores pueden aplicarse?

Ejercicio 3.17

¿Qué relación existe entre el riesgo relativo y la odds ratio?

Número necesario de casos tratados
Definición


El número necesario de pacientes a tratar para evitar un caso (NNT, *number needed to treat/to be treated*) expresa el esfuerzo que hay que realizar para prevenir un caso de enfermedad o la muerte. Se calcula mediante el inverso de la reducción absoluta del riesgo ($1/RA$).

Ejemplo 3.12 (Cont.)


En el **ejemplo 3.12**, si asignamos 1.000 casos a $X+$, cabe esperar que 53 casos desarrollen $Y+$. Si, en cambio, estos mismos 1.000 casos se asignan a $X-$, cabría esperar sólo 9, con un «ahorro» de 44 casos por 1.000 «reasignados». Si se necesitan 1.000 para bajar 44, el número necesario para evitar un caso (NNT) sería 22,7, es decir, 23 casos.

Nota técnica

Esta medida es muy interpretable, pero hay que tener en cuenta que sus propiedades estadísticas son pobres (es muy poco estable de una muestra a otra).

Si en lugar de basarnos en probabilidades o riesgos nos basamos en tasas, estas medidas se definen de forma similar.

Soluciones a los ejercicios

3.1 Dado que $P(B|A)$ debe representar la probabilidad del suceso B dentro del total de casos que cumplen A, se trata de dividir la probabilidad de ser a la vez A y B, es decir $P(A \cap B)$, por la probabilidad de ser A, es decir $P(A)$. En resumen, tiene el mismo numerador que $P(A|B)$, pero cambia el denominador.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Gráficamente sería el cociente entre la zona $A \cap B$ y la zona A

3.2 Sens = $P(+|E)$ Proporción de positivos en el conjunto de enfermos.

Esp = $P(-|S)$ Proporción de negativos en el conjunto de sanos.

VP+ = $P(E|+)$ Proporción de enfermos en el conjunto de positivos.

VP- = $P(S|-)$ Proporción de sanos en el conjunto de negativos.

3.3 Una prueba sería extremadamente específica si, aplicada a un conjunto de sanos, casi el 100% dan negativo. Y un resultado tendrá un elevado valor predictivo negativo si casi el 100% de los que dan negativo están realmente sanos.

3.4 La tabla correspondiente es:

	+	-	TOTAL
Enfermo	132		132
Sano		868	868
TOTAL	132	868	1.000

$$VP+ = P(E|+) = 132 / 132 = 1,00 = 100\%$$

$$VP- = P(S|-) = 868 / 868 = 1,00 = 100\%$$

3.5 Para que $VP+ = 100\%$ se requiere que los casos positivos provengan todos de los enfermos, es decir, que ningún sano dé positivo. Por tanto, se requiere que la especificidad sea del 100% para tener un $VP+$ del 100%. (Similarmente, para que el $VP-$ sea del 100% se requiere que la sensibilidad sea del 100%.)

3.6 La especificidad y el valor predictivo positivo.

3.7 Compare sus respuestas con las definiciones anteriores del inicio del punto 1.4. Discútalas con un colega y vuelva a comprobarlo con las definiciones anteriores.

$$3.8 \text{ Sens} = P(+|E) = 712/1000 = 0,712 = 71,2\%$$

$$\text{Esp} = P(-|S) = 752/1000 = 0,752 = 75,2\%$$

$$VP+ = P(E|+) = 712/960 \approx 0,742 = 74,2\%$$

$$VP- = P(S|-) = 752/1040 \approx 0,723 = 72,3\%$$

$$3.9 \quad P(S|+) = \frac{P(+|S) \cdot P(S)}{P(+|S) \cdot P(S) + P(+|E) \cdot P(E)} = \frac{0,248 \cdot 0,868}{0,248 \cdot 0,868 + 0,712 \cdot 0,132} = 0,696$$

3.10 La tabla correspondiente es:

	AA: Y+	NO AA: Y-	TOTAL
EFV	36	15	51
PI	9	40	49
TOTAL	45	55	100

Riesgo en los expuestos a EFV: $P(Y+|EFV) = 36 / 51 \approx 0,70 = 70\%$

Riesgo en los expuestos a PI: $P(Y+|PI) = 9 / 49 \approx 0,18 = 18\%$

3.11 La probabilidad de sacar un «3» vale $1/6 = 0,167 = 16,7\%$. La *odd* respectiva es $1/5 = 0,2 = 20\%$

3.12 *Odd* (dolor lumbar) = $P(\text{dolor lumbar}) / P(\text{no dolor lumbar}) = 0,5 / 0,5 = 1$ (las *odds* de padecer dolor lumbar a lo largo de la vida están «1 a 1»).

Odd (gripe) = $0,8 / 0,2 = 4$ (las *odds* de gripe están 4 a 1: por cada persona que no padecerá gripe, hay 4 que si la tendrán).

Odd (esclerosis múltiple) = $0,001 / 0,999 = 0,001001001 \approx 0,001$ (las *odds* de esclerosis múltiple están 1 por 1.000).

3.13 Los resultados que figuran a continuación muestran que los riesgos y las *odds* de la última tabla no coinciden con los anteriores. ¿Qué ha pasado? Nótese que la variable respuesta (bronquitis) de la tercera tabla ya no depende de las observaciones, pues, por diseño, se ha dejado fija: se ha elaborado una tabla que tiene, porque así lo hemos querido, la mitad de pacientes con bronquitis y la mitad sin bronquitis. Estos datos, obtenidos de dos muestras de la variable respuesta, ya no sirven para calcular ni los riesgos ni las *odds* de desarrollar una bronquitis. Nótese que en la segunda tabla, en la que había una muestra de fumadores y una de no fumadores, sí que podían calcularse los riesgos y las *odds* condicionados a fumador o a no fumador.

	Riesgos			Odds		
	Tabla 1	Tabla 2	Tabla 3	Tabla 1	Tabla 2	Tabla 3
En los expuestos: X+	0,71	0,71	0,85	2,47	2,47	5,53
En los no expuestos: X-	0,25	0,25	0,42	0,33	0,33	0,74

3.14 Para calcular la RV de un resultado negativo procederemos:

$$RV = \frac{P(-|E)}{P(-|S)} = \frac{1 - 0,712}{0,752} = 0,38$$

Con la misma proporción de enfermos tenemos que:

Odds a posteriori = $RV \cdot \text{Odds a priori} \rightarrow 0,38 \cdot 1/3 \approx 0,13$

3.15 Como cabía esperar, la tercera tabla, en la que se había dejado fijo el número de casos con y sin la enfermedad, no permite calcular los riesgos, ni el relativo ($2,87 \neq 2,00$) ni el atribuible ($0,46 \neq 0,42$). La sorpresa es que sí que permite calcular la *odds ratio*, ya que se obtienen los mismos valores (7,50).

	Riesgos			Odds		
	Tabla 1	Tabla 2	Tabla 3	Tabla 1	Tabla 2	Tabla 3
En los expuestos: X+	0,71	0,71	0,85	2,47	2,47	5,53
En los no expuestos: X-	0,25	0,25	0,42	0,33	0,33	0,74
Riesgo atribuible	0,46	0,46	0,42			
Riesgo relativo	2,87	2,87	2,00			
Odds ratio				7,50	7,50	7,50

3.16 Por lo dicho anteriormente, sólo la *odds ratio*. El riesgo relativo y el riesgo atribuible no tienen valor en este tipo de diseños, ya que su valores dependerán del número de controles que se haya decidido seleccionar para cada caso.

3.17 Son estimadores de la relación entre dos variables dicotómicas, una supuesta respuesta y una supuesta causa. Cuando la proporción de la respuesta es muy pequeña, dan valores muy parecidos. Se interpretan igual.