



Valoración de pruebas diagnósticas

Preparado por Luis M. Molinero (Alce Ingeniería)

[Artículo en formato PDF](#)

CorreoE: bioestadistica@alceingenieria.net

Octubre 2002

www.seh-lelha.org/stat1.htm

Esta página se acompaña con 2 [calculadoras](#) en línea

Aunque la literatura médica y el diseño de estudios de investigación ponen en general un mayor énfasis en el tratamiento de las enfermedades, también las tareas diagnósticas constituyen un núcleo muy importante de la práctica médica. Los procedimientos para evaluar cuantitativamente la eficacia de un proceso de diagnóstico son sobradamente conocidos y se basan en determinar su capacidad para clasificar correctamente los sujetos en dos o más grupos. Se trata de determinar cómo se ajustan a la realidad los resultados de clasificación, obtenidos con la prueba diagnóstica, y para ello es preciso conocer esa realidad mediante algún método alternativo y completamente fiable, lo que en ocasiones no es posible.

El caso más sencillo de prueba diagnóstica es aquél en el que el resultado de la prueba es dicotómico, pudiendo entonces reflejarse en una tabla como la de la figura

		Patología		
		Enfermo E+	Sano E-	
Prueba	Positivo T+	a	b	Positivos Np = a+b
	Negativo T-	c	d	Negativos Nn = c+d
		Total enfermos Ne = a+c	Total sanos Ns = b+d	Total N

En este contexto se utilizan dos índices para evaluar la calidad de la prueba diagnóstica:

- **Sensibilidad**, es la proporción de verdaderos positivos identificados por la prueba del total de enfermos.

$$\text{Sensibilidad} = S = \frac{a}{a+c}$$

- **Especificidad**, es la proporción de verdaderos negativos identificados por la prueba del total de sanos.

$$\text{Especificidad} = E = \frac{d}{b+d}$$

Aunque éstos son dos índices de calidad de la prueba, en la práctica clínica las preguntas a las que interesa responder son: si un sujeto ha resultado positivo, ¿cuál es la probabilidad de que esté verdaderamente enfermo? $P(E+/T+)$, o por el contrario, si el sujeto resultó negativo en la prueba ¿cuál es la probabilidad de que realmente esté sano? $P(E-/T-)$. Estas dos probabilidades se pueden calcular aplicando el [teorema de Bayes](#), siempre que sepamos la probabilidad de que el sujeto esté enfermo antes de realizar la prueba, que se conoce como **probabilidad pre-prueba**. Si no tenemos ninguna información adicional sobre el sujeto, dicha

probabilidad será la **prevalencia** de la patología en la población, aplicable sólo en el caso de programas de cribado o "screening" sobre la población general, ya que en la práctica habitual los sujetos candidatos a una prueba diagnóstica lo son por las sospechas deducidas de la anamnesis o por una sintomatología previa, y por tanto la probabilidad de que padezcan la enfermedad bajo sospecha será superior a la prevalencia de ésta en la población general.

Si llamamos P a la probabilidad pre-prueba de padecer la enfermedad, S a la sensibilidad de la prueba y E a su especificidad, tenemos las siguientes fórmulas para calcular la **probabilidad post-prueba** de que el sujeto esté enfermo cuando resultó positivo o de que esté sano cuando resultó negativo:

$$\Pr(E+ / T+) = \frac{S \cdot P}{S \cdot P + (1 - E) \cdot (1 - P)}$$

$$\Pr(E- / T-) = \frac{E \cdot (1 - P)}{E \cdot (1 - P) + (1 - S) \cdot P}$$

donde la probabilidad condicional la hemos representado de la forma habitual con el símbolo $/$. Así $\Pr(E+/T+)$ significa suceso $E+$ (enfermo) condicionado a que haya ocurrido el suceso $T+$ (resultado + en la prueba).

Si calculamos estas probabilidades únicamente con los datos de nuestra tabla, la primera de ellas $\Pr(E+/T+)$ corresponde a la proporción de sujetos que verdaderamente tienen la enfermedad, de entre los que dieron positivo, y se conoce como **valor predictivo positivo** $VP+$

$$VP+ = \frac{a}{a + b}$$

Igualmente podemos calcular en la tabla la proporción de sujetos verdaderamente sanos sobre el total de los que dieron negativo, **valor predictivo negativo** $VP-$

$$VP- = \frac{d}{c + d}$$

Puesto que esos dos índices $VP+$ y $VP-$ son los que interesan en la práctica clínica, parecería natural utilizarlos como índices de comparación a la hora de evaluar dos métodos diagnósticos diferentes. Sin embargo presentan un grave inconveniente, ya que si se calculan a partir de la tabla dependen de la proporción de enfermos en la muestra estudiada. Por ello para una determinada prueba resulta necesario determinar unos índices de valoración que, respondiendo a las necesidades reales en cuanto a la clasificación de pacientes, sin embargo no dependan de esa proporción de enfermos en la muestra.

Así, podemos calcular el cociente entre la tasa de verdaderos positivos y la tasa de falsos positivos:

$$CP+ = \frac{\Pr(T+ / Enfermos)}{\Pr(T+ / Sanos)}$$

Este cociente $CP+$ se conoce como **cociente de probabilidad positivo** o también como cociente de verosimilitud (likelihood ratio of positive test). Valores mayores de $CP+$ indican mejor capacidad para diagnosticar la presencia de enfermedad.

De la misma forma, podemos definir un **cociente de probabilidad negativo** $CP-$

$$CP- = \frac{\Pr(T- | Enfermos)}{\Pr(T- | Sanos)}$$

Vemos que valores de $CP-$ menores indican una mejor capacidad diagnóstica de la prueba.

Se puede determinar unas fórmulas para expresar los cocientes de probabilidad en función de la sensibilidad S y de la especificidad E

$$CP+ = \frac{S}{1-E}$$

$$CP- = \frac{1-S}{E}$$

La ventaja de los cociente $CP+$ y $CP-$ frente a los valores predictivos positivo y negativo de la prueba radica en que, a diferencia de éstos, no dependen de la proporción de enfermos en la muestra, sino tan sólo de la sensibilidad y especificidad de ésta, de ahí su utilidad a la hora de comparar pruebas diagnósticas. Además si conocemos o podemos hacer una estimación de la probabilidad pre-prueba de que un sujeto padezca la enfermedad, utilizando los cocientes de probabilidad, al realizar el test podemos "corregir" ese valor de acuerdo con el resultado, de tal manera que la probabilidad aumenta o disminuye según que el resultado sea positivo o negativo, aplicando la siguiente fórmula

$$P_{post} = \frac{P \cdot CP}{1 + P \cdot (CP - 1)}$$

donde P es la probabilidad pre-prueba, CP el correspondiente cociente de probabilidad (positivo si deseamos calcular la probabilidad de que padezca la enfermedad, negativo en caso contrario) y P_{post} es la probabilidad post-prueba.

Veamos un ejemplo de todo esto con números:

		Patología		
		Enfermo E+	Sano E-	
Prueba	Positivo T+	72	100	Positivos 172
	Negativo T-	18	150	Negativos 168
		Total enfermos 90	Total sanos 250	Total 340

Aplicando las fórmulas anteriores tenemos

	Valor estimado	Intervalo de confianza 95 %
Sensibilidad	0.8	0.72 : 0.88
Especificidad	0.6	0.54 : 0.66
CP+	2	1.66 : 2.4
CP-	0.33	0.22 : 0.51

Evidentemente los valores de los diferentes índices son estimaciones realizadas mediante un experimento diseñado al efecto, por lo que es necesario calcular algún indicador de su grado de incertidumbre, como puede

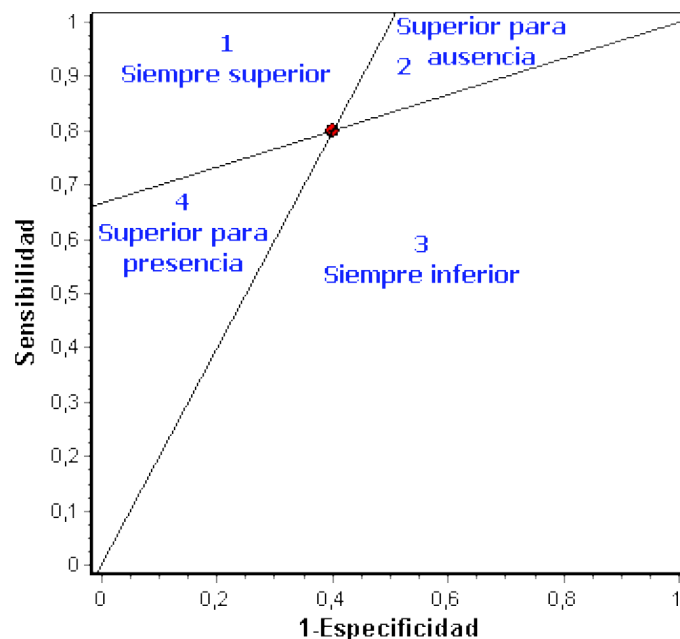
ser un intervalo de confianza del 95%. Para el caso de la sensibilidad y especificidad es sencillo ya que se trata de dos proporciones. En el caso de los cocientes de probabilidad es algo más complejo y utilizamos una aproximación.

Para este ejemplo, si la probabilidad pre-prueba de un sujeto es de 0.7 y el resultado de la prueba resulta positivo la probabilidad de que padezca la enfermedad se modifica, aplicando la fórmula descrita anteriormente, a 0.82, mientras que si fue negativo se modifica a 0.44.

Si comparamos dos pruebas diagnósticas A y B, y calculamos sus cocientes de probabilidad positivos, y vemos que $CP_{+A} > CP_{+B}$, diremos que la prueba A es mejor que la B para confirmar la presencia de enfermedad.

Si en cuanto a los cocientes de probabilidad negativos vemos que $CP_{-A} < CP_{-B}$, diremos que la prueba A es mejor que la B para confirmar la ausencia de enfermedad.

Comparando los cocientes de probabilidad de dos pruebas diagnósticas existen cuatro posibles resultados que podemos reflejar en una gráfica. En el eje de las Y representamos la *sensibilidad*, en el eje de las X se representa ($1 - \text{Especificidad}$). El punto representa los valores de Sensibilidad y ($1 - \text{Especificidad}$) observados para nuestra prueba A. Trazamos una línea desde el origen (0,0) que pase por el punto A y otra desde el valor (1,1) que también pase por ese punto, quedando dividido el gráfico en cuatro zonas como vemos en la figura, de tal manera que cualquier otra prueba cuyo resultado representemos en esa gráfica estará en alguna de las cuatro zonas

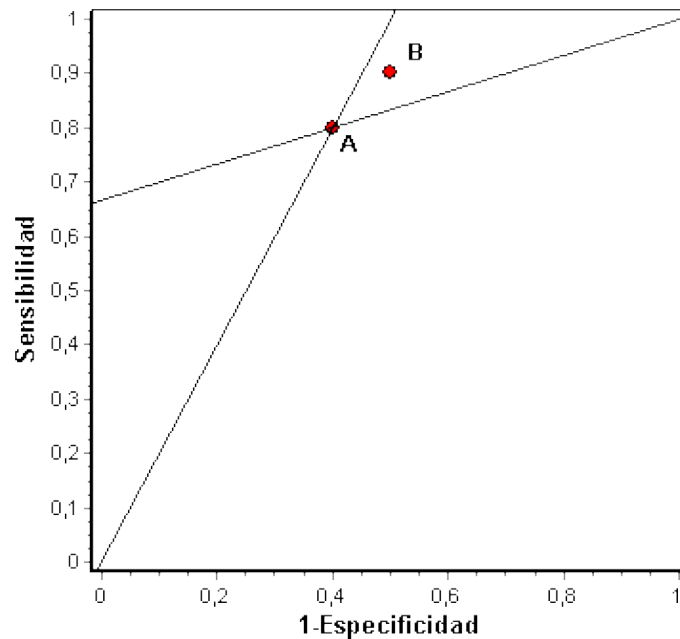


Las zonas corresponden a las siguientes situaciones:

1. En esta región la prueba B es siempre superior a la A, tanto para confirmar presencia como ausencia de enfermedad
2. Aquí la prueba B sólo es superior a la A para confirmar ausencia de enfermedad
3. En esta zona B siempre es inferior a A

4. Si el resultado de B está en este área, será superior a A para confirmar la presencia de enfermedad

Supongamos que deseamos comparar la prueba A con otra prueba B cuya sensibilidad es 0.9 y especificidad 0.5. Representándolo en la figura obtenemos

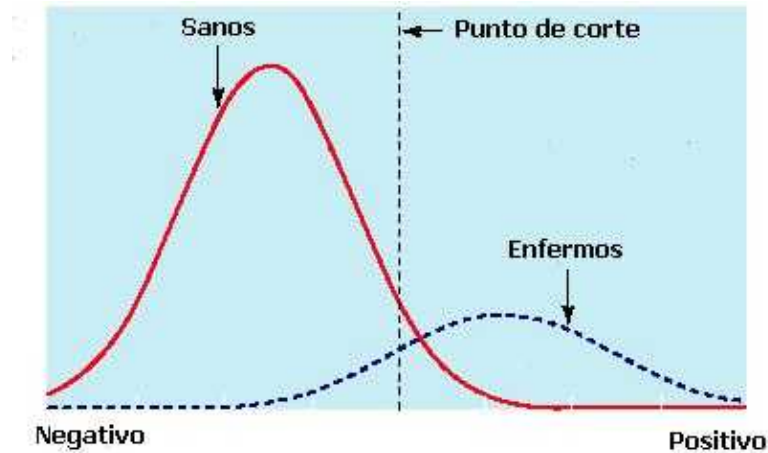


Donde vemos que B cae en la zona 2 por lo que es una prueba mejor que la A para confirmar la ausencia de enfermedad, como podemos determinar también numéricamente al calcular los correspondientes cocientes de probabilidad $CP+ = 1.8$ y $CP- = 0.2$

Todas estas consideraciones se refieren únicamente a la capacidad discriminadora de una prueba. La decisión sobre qué prueba diagnóstica es más adecuada, es siempre algo más complicado, puesto que intervienen otros aspectos, como son el coste de la prueba, los riesgos que supone para el paciente, valorar las consecuencias que conlleva un falso positivo o un falso negativo...

Evidentemente para comparar los parámetros de dos pruebas diagnósticas habrá que considerar que los valores obtenidos son sólo estimaciones y están sometidos por tanto a posibles errores de muestreo, por lo que habrá que efectuar el correspondiente contraste estadístico para determinar si las diferencias encontradas son suficientemente importantes como para no poder ser atribuidas al azar.

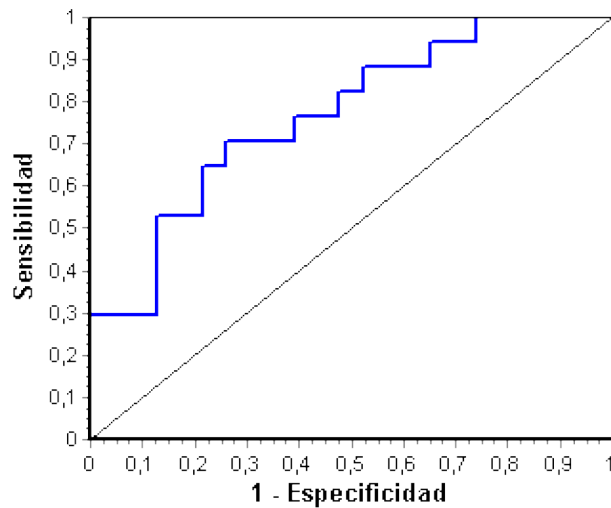
Hasta ahora hemos estudiado el caso de un resultado dicotómico, pero en muchas ocasiones las pruebas diagnósticas son cuantitativas, sobre todo cuando corresponden a determinaciones analíticas. Evidentemente podemos utilizar todas las consideraciones hechas hasta el momento si fijamos un punto de corte, un valor determinado de la prueba, que marque el límite entre sano y enfermo. Pero eso no suele ser una tarea sencilla. Podemos reflejarlo esquemáticamente en una figura:



Existe una zona de posibles resultados de la prueba para la que la distribución de sujetos sanos y enfermos se solapan. Si queremos aumentar la probabilidad de detectar pacientes enfermos, en la gráfica corresponde a mover el punto de corte hacia la izquierda, vemos que entonces también aumentaremos el número de falsos positivos. Si movemos el punto de corte hacia la derecha, disminuirémos los falsos positivos, pero a costa de aumentar el de falsos negativos. Resumiendo, un aumento de la sensibilidad disminuye la especificidad, y viceversa.

Una herramienta útil para evaluar la capacidad diagnóstica de una prueba cuantitativa para todos los posibles puntos de corte, es la denominada **curva ROC**. También nos servirá para comparar diferentes pruebas.

Para obtener la curva ROC, se calcula la sensibilidad y especificidad para cada uno de los diferentes valores observados en nuestros datos y se representan en una gráfica, similar a la anteriormente descrita, con la *Sensibilidad* en el eje de las Y, ($1 - \text{Especificidad}$) en el eje de las X.



En la situación ideal, una prueba que discrimina perfectamente, quedaría representado en la gráfica como una línea que coincidiría con los lados izquierdo y superior del cuadrado; mientras que una prueba que no discrimine en absoluto, corresponde a la línea diagonal (a 45°) que aparece en la figura. Por lo tanto, cuanto más desplazada esté la curva ROC hacia el vértice superior izquierdo, mejor es la capacidad discriminatoria de la prueba. Precisamente una forma de evaluar de manera global esa capacidad de discriminación consiste en calcular el área del polígono que queda debajo de la curva ROC, y se denomina **área bajo la curva**,

sirviendo como índice de comparación entre pruebas diagnósticas, cuanto mayor es el área mejor es la capacidad diagnóstica.

Calculadoras

- [Calculadora a partir de los datos de una tabla 2x2](#)
- [Calculadora de la probabilidad post-prueba conociendo la sensibilidad y la especificidad](#)

Enlaces relacionados

- [Conceptos básicos para interpretar los resultados de un estudio sobre pruebas diagnósticas](#)
Web de la Asociación Española de Atención Primaria
- [Calculadora para pruebas diagnósticas de la red CASPE \(fichero Excel\)](#)
- *Statistics Notes: Diagnostic tests 1: sensitivity and specificity*
D G Altman and J M Bland
BMJ 1994; 308: 1552. [\[Full text\]](#)
- *Statistics Notes: Diagnostic tests 2: predictive values*
Douglas G Altman and J Martin Bland
BMJ 1994; 309: 102. [\[Full text\]](#)
- *Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots*
D G Altman and J M Bland
BMJ 1994; 309: 188. [\[Full text\]](#)



[Índice de artículos](#)

[Principio de la página](#) ▲