



Medicina de Familia SEMERGEN

www.elsevier.es/semergen



FORMACIÓN CONTINUADA - METODOLOGÍA Y TÉCNICAS

La curva ROC

J.A. Martínez Pérez^{a,*} y P.S. Pérez Martín^b

^a Miembro de la Comisión Nacional de Calidad de SEMERGEN, Madrid, España

^b Médico de familia, Centro de Salud Palma Norte, Madrid, España

Recibido el 26 de abril de 2022; aceptado el 15 de julio de 2022



PALABRAS CLAVE

Curva ROC;
Punto de corte;
Capacidad
discriminativa

Resumen La curva ROC es una herramienta estadística que se utiliza para evaluar la capacidad discriminativa de una prueba diagnóstica dicotómica. Se trata de curvas en las que se presenta la sensibilidad en función de los falsos positivos (complementario de la especificidad) para distintos puntos de corte. Son útiles para elegir el punto de corte más adecuado de una prueba, conocer el rendimiento global de esta y comparar la capacidad discriminativa de 2 o más pruebas diagnósticas.

© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

ROC curve;
Cut-off point;
Discriminative
capacity

ROC curve

Abstract The ROC curve is a statistical tool used to evaluate the discriminative capacity of a dichotomous diagnostic test. These are curves in which sensitivity is presented as a function of false positives (complementary to specificity) for different cut-off points. They are useful for choosing the most appropriate cut-off point for a test, knowing its overall performance and comparing the discriminative capacity of 2 or more diagnostic tests.

© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Published by Elsevier España, S.L.U. All rights reserved.

Introducción

El análisis con base en curvas *receiver operating characteristic* (ROC, «característica operativa del receptor», aunque

la traducción quizá se ajuste más a «curva de eficacia diagnóstica») constituye un método estadístico para determinar la exactitud diagnóstica de test que utilizan escalas continuas, siendo utilizadas con 3 propósitos específicos: determinar el punto de corte en el que se alcanza la sensibilidad y especificidad más altas, evaluar la capacidad discriminativa del test diagnóstico y comparar la capacidad discriminativa de 2 o más test diagnósticos que expresan sus resultados como escalas continuas¹.

* Autor para correspondencia.

Correo electrónico: jmartinezp@semergen.es
(J.A. Martínez Pérez).

<https://doi.org/10.1016/j.semerg.2022.101821>

1138-3593/© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Tabla 1 Capacidad discriminativa del cuestionario SCOFF para la detección de trastornos de la conducta alimentaria en Atención Primaria

Umbral de detección de respuestas positivas	Sensibilidad, %	Especificidad, %
0+	100 (73,5-100)	0 (0-1,4)
1+	91,6 (61,5-99,7)	29,7 (24,2-35,2)
2+	91,6 (71,8-100)	91 (87,5-94,5)
3+	41,6 (15,1-72,3)	97 (94-99,3)
4+	8,3 (0,2-38,4)	99,3 (96,2-99,5)
5+	0 (0-26,4)	99,7 (99,2-99,9)

Los valores se presentan con un intervalo de confianza del 95%.

Estas curvas fueron desarrolladas en el seno de las investigaciones militares estadounidenses. Su primera aplicación fue la detección de señales de radar durante los años 50 del siglo pasado. En los 60, Green y Swets² la usaron para experimentos psicofísicos y más tarde, en los 70, el radiólogo Leo Lusted³ las empleó para la toma de decisiones en la interpretación de imágenes radiológicas.

Cuando los valores de una prueba diagnóstica siguen una escala cuantitativa, la sensibilidad y la especificidad varían según el punto de corte elegido para clasificar a la población como enferma o no enferma. Mediante el uso de las curvas ROC hacemos una medición global del conjunto de todos los puntos de corte y precisamos cuál de ellos alcanza la sensibilidad y la especificidad más altas⁴, considerándose «positivos» aquellos valores mayores o iguales al punto de corte y «negativos» aquellos valores menores al mismo.

Para construir esta curva es necesario calcular la sensibilidad y la especificidad para todos los posibles puntos de corte⁴. En la tabla 1 exponemos un ejemplo basado en un reciente trabajo nuestro, sobre la validez del cuestionario SCOFF para la detección de trastornos de la conducta alimentaria⁵. Como podemos observar, la mayor sensibilidad y especificidad conjunta concurren en el punto de corte 2.

La sensibilidad o proporción de verdaderos positivos se sitúa en el eje de ordenadas y el complemento de la especificidad (1 – especificidad) o proporción de falsos positivos, en el de abscisas⁴. Ambos ejes engloban valores entre 0 y 1 (0 a 100%). La curva ROC se dibuja uniendo los pares de valores (sensibilidad y complemento de la especificidad) correspondientes a cada punto de corte (fig. 1).

Punto de corte que determina la sensibilidad y la especificidad más altas

Es el que presenta el mayor índice de Youden, calculado según la fórmula (sensibilidad + especificidad – 1). Gráficamente, este corresponde al punto de la curva ROC más cercano al ángulo superior-izquierdo del gráfico (punto 0,1), es decir, más cercano al punto del gráfico cuya sensibilidad y especificidad son del 100%.

Sin embargo, en este aspecto hay que considerar que este índice de Youden indica cuál es el punto de corte que

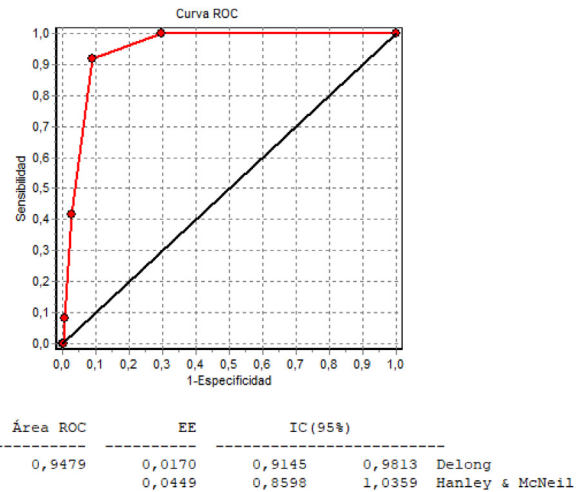


Figura 1 Curva ROC. Capacidad discriminativa del cuestionario SCOFF.

determina la sensibilidad y especificidad más altas conjuntamente, pero no siempre este punto de corte determina la sensibilidad ni la especificidad más altas del test.

Generalmente, la sensibilidad más alta corresponde a un punto de corte, mientras que la especificidad más alta corresponde a otro, y en ocasiones interesa disponer de un test diagnóstico altamente sensible o altamente específico. En este caso lo más adecuado es, una vez conocidas la sensibilidad y la especificidad de los diferentes puntos de corte, decantarse por aquel que tenga la mayor sensibilidad o especificidad, según sea el objetivo que nos propongamos¹.

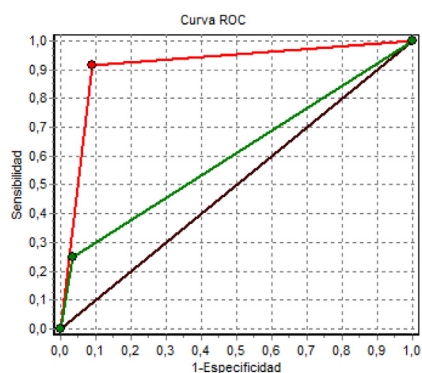
Evaluar la capacidad discriminativa de un test diagnóstico

En la curva ROC existe un área denominada área bajo la curva (ABC, o AUC en terminología anglosajona, por *area under the curve*) que mide la facultad discriminativa del test, es decir, su capacidad de diferenciar sujetos sanos versus enfermos. El AUC la podemos definir como la probabilidad de clasificar correctamente a un par de individuos (uno sano y otro enfermo) seleccionados al azar al aplicarles la prueba⁴.

El AUC refleja cómo de bueno es el test que queremos utilizar para discriminar pacientes con y sin enfermedad a lo largo de todo el rango de puntos de corte posibles. Una prueba con discriminación perfecta (sensibilidad igual a 1 y especificidad igual a 1) estaría representada por una curva ROC situada en el lado izquierdo y superior del gráfico. Si la curva no fuera discriminativa (sensibilidad y especificidad igual a 0,5), estaría representada por la diagonal del gráfico.

Según el área que obtengamos, podemos estimar cuál es el grado de discriminación del test. Cuanto más se acerque a 1, más discriminativo será, y cuanto más se aproxime a 0,5, menos discriminativo será. Una curva con un valor de 0,75 consideraremos que tiene una capacidad discriminativa media.

A título orientativo, podemos usar los siguientes intervalos para interpretar los valores de AUC⁶:



Curva	Área ROC	EE (DeLong)	IC (95%)	
1	0,9135	0,0425	0,8302	0,9969
2	0,6071	0,0655	0,4787	0,7355

Prueba de homogeneidad de áreas		
Ji-cuadrado	gl	Valor p
15,3907	1	0,0001

Figura 2 Comparación de curvas ROC de los cuestionarios SCOFF y EAT-26.

- 0,5: el modelo no tiene capacidad discriminativa.
- 0,5-0,6: test malo.
- 0,6-0,75: test regular.
- 0,75-0,9: test bueno.
- 0,9-0,97: test muy bueno.
- 0,97-1: test excelente.

Por otra parte, hay que considerar que el AUC es un estimador muestral de una población; por este motivo es necesario hallar el intervalo del mismo con una confianza del 95%. Si este intervalo incluye el 0,50% (por ejemplo, 0,43-0,68%) no podemos considerar que el test que evaluamos sea discriminativo¹.

Comparar la capacidad discriminativa de 2 o más test diagnósticos

Para lograr este fin es necesario, en primer lugar, comprobar que se cumplen los siguientes requisitos: los test a comparar deben ser medidos simultáneamente, aplicados sobre los mismos sujetos y contrastados con el mismo patrón oro.

La comparación consiste en confrontar las respectivas AUC de cada test, siendo más discriminativo el que tenga

una mayor AUC. Sin embargo, para poder afirmar que un test tiene superior capacidad discriminativa sobre otro, es preciso comparar estadísticamente ambas AUC según los métodos descritos por Hanley y McNeil⁷ o DeLong et al.⁸ (preferentemente este último). Si no podemos utilizar estas herramientas, solamente podremos sugerir que un test diagnóstico es más discriminativo que otro, pero no afirmarlo.

Como podemos comprobar en la figura 2, el AUC del test SCOFF (0,9135) podemos considerarla muy buena y es superior a la del test EAT-26 (0,6071), que definiremos como regular. Con base en estos resultados, sugeriremos que el primer test es más discriminativo que el segundo, pero además, al comparar ambas áreas con el método DeLong, vemos que el primero es estadísticamente significativo respecto al segundo ($p=0,0001$), lo que nos lleva a afirmarlo.

Por otra parte, también podemos observar que el intervalo de confianza del test EAT-26 está comprendido entre 0,4787 y 0,7355%, es decir, incluye el 0,50%, lo que nos conduce a concluir que este test en nuestro estudio no era discriminativo.

Limitaciones de su uso: estas curvas solo contemplan 2 estados clínicos posibles (sano, enfermo) y no sirven para situaciones en que se trate de discernir entre más de 2 enfermedades.

Bibliografía

1. Cerda J, Cifuentes L. Uso de curvas ROC en investigación clínica. Aspectos teórico-prácticos. Rev Chil Infect. 2012;29:138-41, <http://dx.doi.org/10.4067/S0716-10182012000200003>.
2. Green DM, Swets JA. Signal detection theory and psychophysics. New York: John Wiley & Sons, Inc.; 1966.
3. Lusted LB. Signal detectability and medica decision-making. Science. 1971;171:1271-9.
4. García de Yébenes Pous MJ, Rodríguez Salvanés F, Carmona Ortelis L. Validación de cuestionarios. Reumatol Clin. 2009;5:171-7.
5. Pérez Martín PS, Martínez Pérez JA, Espejo González A. Is early detection of eating disorders possible in primary care? Fam Med Prim Care Rev. 2022;24:61-5, <http://dx.doi.org/10.5114/fmpcr.2022.113015>.
6. Wikipedia [Internet]. Curva ROC [consultado 29 Jun 2022]. Disponible en: https://es.wikipedia.org/wiki/Curva_ROC.
7. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148:839-43.
8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics. 1988;44:837-45.