Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones.

Ana Rocío del Valle Benavides

Tutor: Juan Manuel Muñoz Pichardo
Trabajo Final de Grado en Matemáticas
Departamento de Estadística e Investigación Operativa

Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones

Ana Rocío del Valle Benavides

Agradecimientos

A *Juan Manuel Muñoz Pichardo* por su dedicación, correcciones y largas tutorías. Al equipo *Môlab* (Mathematical Oncology Laboratory) por incluirme en sus labores de investigación del Glioblastoma, dándome pié a conocer así las curvas ROC y sus aplicaciones.

Tutor: Juan Manuel Muñoz Pichardo.

Universidad de Sevilla.
Departamento de Estadística e Investigación Operativa.

Índice general

Resumen

La curva ROC es una herramienta estadística utilizada en el análisis de la clasificar la capacidad discriminante de una prueba diagnóstica dicotómica. Es decir, una prueba, basada en una variable de decisión, cuyo objetivo es clasificar a los individuos de una población en dos grupos: uno que presente un evento de interés y otro que no. Esta capacidad discriminante está sujeta al *valor umbral* elegido de entre todos los posibles resultados de la *variable de decisión*, es decir, la variable por cuyo resultado se clasifica a cada individuo en un grupo u otro. La curva es el gráfico resultante de representar, para cada valor umbral, las medidas de *sensibilidad* y *especificidad* de la prueba diagnóstica. Por un lado, la sensibilidad cuantifica la proporción de individuos que presenta el evento de interés y que son clasificados por la prueba como portadores de dicho evento. Por otro lado, la especificidad cuantifica la proporción de individuos que no lo presentan y son clasificados por la prueba como tal.

Desde su invención en el seno de las investigaciones militares estadounidenses ha formado parte del Análisis Discriminante y la Teoría de la Detección de Señales. Su primera aplicación fue en detección de señales de radar durante los años 50'. En los 60' Green y Swets[1] la utilizaron para experimentos psicofísicos y más tarde, en los 70', el radiólogo Leo Lusted[26] las usó para decisión de disgnóstico mediante imágenes médicas. A partir de entonces, numerosos investigadores han utilizado ésta herramienta en el campo de la sanidad, la economía, la meteorología y más recientemente en el aprendizaje automático.

Abstract

The COR curve is a statistic tool wich is used in sort analysis for sorting out the discriminating ability of a diagnosis dichotomic test. That is to say, a test, based on a decision variable wich aim is to classify the population´s individuals into two groups: one that presents an event and another that doesn't. This discriminant capacity is subject to the *threshold value* chosen from among all possible outcomes of the *decision variable*, ie, the variable by which each individual is classiffied in one group or another. The curve is the graph resulting from representing, for each threshold value, the measures of *sensitivity* and *specificity*. On the one hand, the sensitivity quantifies the proportion of individuals in the sample who present the event and are classified by the test as carriers the mentioned event. On the other hand, specificity quantifies the proportion of individuals who don't present the event and are classified by as such.

Since its invention by the US military´s investigations it has been part of the Discriminant Analysis and Signal Detection Theory. Its first application was for the detection of radar signals during the 50's. In the 60's, Green and Swets[1] used it for psychophysical experiments and later on, in the 70's, the radiologist Leo Lusted[26] used them for diagnosis decisions using medical images. Since then, several authors have used this tool in the sanitary field, such as in economy, meteorology and more recently in machine learning.

Capítulo 1

Introducción

Fueron ingenieros militares estadounidenses, durante la Segunda Guerra Mundial, los que empezaron a utilizar la curva ROC. Las usaron como herramienta para discernir en sus radares qué era una señal de ruido y qué era una señal de ofensiva militar de otro país (torpedos, misiles y similares).

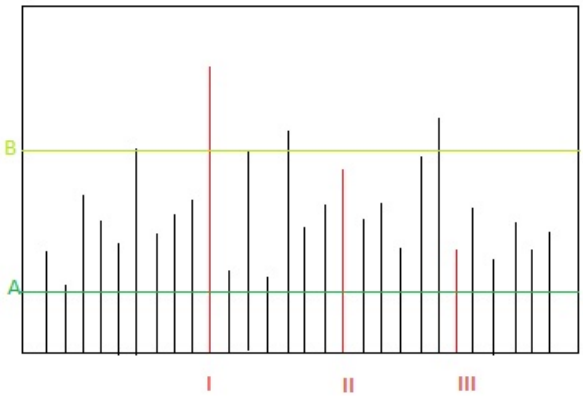
El evento que suscitó éstas investigaciones fué el ataque a la base naval americana de Pearl Harbor (Isla de Oahu, Hawái). Esta ofensiva inesperada fue llevada a cabo por la Armada Imperial Japonesa en la mañana de domingo del 7 de diciembre de 1941. Según el ejército japones, el objetivo fue preventivo, ya que con este ataque querían evitar la intervención de EEUU en sus maniobras militares contra las posesiones ultramarinas de varios de sus países enemigos en el Sudeste Asiático. Lo que sigue es un fragmento de la carta que el almirante Yamamoto envió al capitán Genda, miembro de la Primera División Aérea de la Armada Imperial Japonesa, en busca de su aprobación como piloto para organizar el ataque.

> *Si Japón y Estados Unidos fueran a la guerra, tendríamos que recurrir a una táctica radical. Deberíamos intentar, con toda la fuerza de nuestras Primera y Segunda Divisiones Aéreas, asestar un golpe a la flota estadounidense en Hawái, de forma, que durante un tiempo, Estados Unidos no pudiera avanzar hacia el Pacífico occidental.*(Febrero de 1941)[4].

Al día siguiente del ataque, el 8 de diciembre de 1941, EEUU le declaró la guerra a Japón y consolidó su alianza, hasta entonces soterrada, con Reino Unido. Por su parte Alemania e Italia le declararon la guerra a EEUU en respuesta a la preparación de ofensivas contra Japón.

A partir de este suceso, el ejército norteamericano desarrolló un proyecto que trataba de abordar el siguiente problema:

Figura 1.0.1: Ejemplo de receptor de señal en radar.



La figura 1.0.1 representa la pantalla de un receptor de señal radar, las señales I,II y III corresponden a misiles, sin embargo el resto no y, por tanto, es ruido. Ante una nueva lectura deberíamos fijar un valor umbral a partir de cual considerar que esa señal corresponderá a un misil y por debajo no. Dicho valor umbral, según se escoja, nos puede dar muchos avisos falsos de misiles (*falsos positivos*) o, por el contrario, puede no avisar ante señales que correspondían a misiles (*falsos negativos*).

Si sobre la lectura de la imagen escogemos el valor umbral A nos dará aviso de los tres misiles (*verdaderos positivos*) a la par de que nos dará muchas falsas alarmas, es decir, falsos positivos. Por otro lado, si escogemos el valor umbral B tan sólo nos avisará de uno de los tres misiles, es decir, tendremos sólo un verdadero positivo y dos falsos negativos. Bajo umbral B, también tendremos ruido, es decir falsos positivos. Por lo que parece que lo único bueno del valor umbral B es que tiene muchas, en proporción, señales de ruido bien catalogadas (*verdaderos negativos*).

Así, la finalidad de la curva ROC aportarnos información suficiente para escoger el mejor valor umbral. Además, nos aporta herramientas para, no sólo minimizar los $falsos\ positivos$ y los $falsos\ negativos$, sino para escoger de qué preferimos tener menos: falsas alarmas de misiles o, por el contrario, el número de ellos que no son detectados.

Este mismo razonamiento es el que aplica el radiólogo cuando, ante una imagen médica, trata de decidir si cierta mancha es ruido o alguna anormalidad del tejido en estudio. Así mismo, el médico, cuando trata de validar una prueba para detectar cierta enfermedad, también encuentra la disyuntiva de qué valor umbral escoger y tener o más falsos positivos (*medicina preventiva*) o más falsos negativos (*tamizaje de enfermedades*).

Las ventajas del uso de la curva ROC según Zhou et. al son: que proporciona una representación de la sensibilidad y especificidad para cada valor umbral, que es invariante mediante transformaciones monótonas a los datos de la variable de decisión y que permite comparar dos o más clasificadores en función de su capacidad discriminante.

Tras conocer el evento histórico que dio pié al desarrollo de esta herramienta, el objetivo del trabajo es estudiar la construcción y posterior aplicación de la curva ROC. Para ello, explicaremos los conceptos básicos en los que se basa, a saber, resultados acertados y no acertados, positivos y negativos. Veremos formas de construcción mediante proporciones de resultados de la prueba sobre la muestra (métodos no paramétricos) o ajustando los mismos a alguna distribución conocida (métodos paramétricos). Conoceremos medidas para cuantificar la bondad de un clasificador para un diagnóstico y evaluaremos diferentes formas de escoger el valor umbral o punto de corte.

A continuación, aplicaremos la teoría recogida en la memoria a un estudio real a través de los softwares estadísticos SPSS y R. Usaremos datos reales de una muestra de pacientes a los que se le midieron dos biomarcadores con objetivo de discriminar entre los que paceden y no cáncer de páncreas. Construiremos las curvas ROC para ambas pruebas y determinaremos aquella que clasifica mejor a los individuos.

Finalmente, para ilustrar la utilidad de la técnica objeto de la presente memoria haremos un recorrido por diferentes aplicaciones de las curvas ROC en los campos de la medicina, psicología, economía, meteorología y aprendizaje automático, explicando con qué objetivo artículos, tesis y proyectos de investigación hacen uso de esta herramienta estadística.

Capítulo 2

Construcción de la curva ROC

2.1. Conceptos previos

Sea la *variable aleatoria* $D =' estado'$, que sigue una *distribución Bernoulli* de parámetro p que llamaremos *prevalencia* del evento sobre la población. Dicha variable toma los valores:

$$D \sim Be(p) \Rightarrow D = \begin{cases} 0 & Cuando\, el\, individuo\, no\, presenta\, el\, evento \\ 1 & Cuando\, el\, individuo\, sí\, presenta\, el\, evento \end{cases}$$

Por tanto,

$$prevalencia = p = P_r('estado' = 1) = P_r(D = 1) = P_r(presentar\, el\, evento)$$

y puesto que ambos estados forman un espacio de sucesos:

$$1 - p = P_r('estado' = 0) = P_r(D = 0) = P_r(no\, presentar\, el\, evento)$$

.
Nuestros elementos de partida van a ser:

- Una *muestra aleatoria simple* de un grupo control que no presente el evento de interés, a los que llamaremos *sanos y* una *muestra aleatoria simple* de un grupo que sí lo presente, a los que llamaremos *enfermos* ó, directamente, una $m.a.s$ de la población.
- Una *variable aleatoria* x que mida cierta característica en cada individuo, cuyo resultado puede ser continuo ó discreto.
- Una variable aleatoria Bernoulli que llamaremos $y =' prueba'$ de la que queremos estudiar su eficiencia discriminante, que tomará dos resultados:

positivo ó negativo, en función del valor de x respecto al que denominaremos *punto de corte* o *valor umbral* c. Este será cada uno de los posibles valores que toma la variable x.

$$'Prueba' = \begin{cases} Positivo \equiv y = 1 & si\, x \geq c \\ Negativo \equiv y = 0 & si\, x < c \end{cases}$$

Observación. Es equivalente tener una muestra de la población y conocer la variable $'estado'$ de cada individuo, a tener una muestra de un grupo que es seguro que presenta el evento y otra muestra de otro grupo que no lo presenta. Sin embargo, a efectos de la *prevalencia* no.

Si extraemos una muestra directamente de la población, un estimador de la prevalencia sería la razón:

$$p = \frac{número\, de\, enfermos\, de\, la\, muestra}{cantidad\, de\, individuos\, de\, la\, muestra}$$

Pero si componemos nuestra muestra con una extracción de un grupo de sanos y otra extracción de un grupo de enfermos dicha razón no lo será. Por tanto deberemos, es este caso, extraer tántos enfermos, en función de sanos, como nos indique la prevalencia, usando un estimador de ella proporcionada por un experto en caso de desconocerla.

Lo siguiente es aplicar nuestra prueba a sanos y enfermos. Dicha prueba nos dividirá a la población en cuatro subgrupos como presenta la siguiente tabla de contingencia:

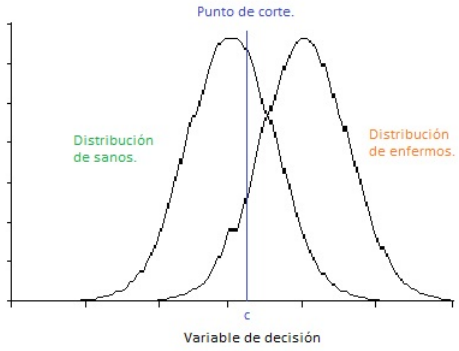
	$Enfermo \equiv D = 1$	$Sano \equiv D = 0$
$Prueba+ \equiv y = 1$	Verdadero positivo(V_+)	Falso positivo(F_+)
$Prueba- \equiv y = 0$	Falso negativo(F_-)	Verdadero negativo(V_-)

De manera equivalente, un individuo será:

$$\begin{cases} V_+ & si\, (D = 1,\, y = 1) \\ F_+ & si\, (D = 0,\, y = 1) \\ F_- & si\, (D = 1,\, y = 0) \\ V_- & si\, (D = 0,\, y = 0) \end{cases}$$

Supongamos que el resultado de la prueba es una variable aleatoria continua, ésta no tiene porqué tener la misma distribución en el grupo de enfermos que en el de sanos, y en el caso de tenerla no tiene porqué ser bajo los mismos parámetros. De hecho, en el caso límite en el que la distribución y los parámetros fuesen iguales significaría que enfermos y sanos se comportarían igual, es decir,

Figura 2.1.1: Distribuciones solapadas de sanos y enfermos.



que la prueba sería inutil para detectar la enfermedad en cuestión.
En la figura 2.1.1 se representa en el eje de abcisas el valor de la variable x y en el eje de ordenadas un ejemplo de la distribución que podría tener en ambos grupos. El valor c es el punto de corte o valor umbral, por encima del cual nuestra prueba considerará al individuo enfermo aunque su estado real sea sano y viceversa.

Ante ésta gráfica cabe esperar que las distribuciones estén lo más alejadas posibles para que se produzca cuanto menos solapamiento mejor, es decir, falsos negativos y falsos positivos. Ésta separación o no dependerá de la capacidad discriminatoria de la prueba. Por otro lado, respecto a la elección del punto de corte hablaremos más adelante en el Capítulo 4 pero está claro que una vez fijado necesitaremos cuantificar los resultados acertados y los que no. Para ello tenemos los conceptos de *sensibilidad* y *especificidad*.

Sensibilidad: Es la probabilidad de, dado un individuo enfermo, que la prueba lo clasifique como enfermo. También se le denomina valor predictivo positivo:

$$S = P_r(y = 1|D = 1) = \frac{P_r(y = 1 \cap D = 1)}{P_r(D = 1)}$$

Dada su anterior definición poblacional, podemos aproximarla mediante un estimador muestral. Aplicando el axioma de Kolmogorov de « la probabilidad de un suceso es el número de casos favorables entre el número de casos posibles» obtenemos la fórmula:

$$S = \frac{V_+}{V_+ + F_-}$$

O bien, *fracción de verdaderos positivos(FVP)*, exactitud positiva o *recall*. Así mismo, aplicando «la probabilidad de un suceso es igual a 1 menos

la probabilidad de su complementario» tenemos que $1 - S = \frac{V_+ + F_-}{V_+ + F_-} - \frac{V_+}{V_+ + F_-} = \frac{F_-}{V_+ + F_-}$, conocida como la *fracción de falsos negativos(FFN).*

Observación. Si $S = 1 \Rightarrow F_- = 0$.

Especificidad: Es la probabilidad de, dado un individuo sano, que la prueba lo clasifique como sano. También se le denomina valor predictivo negativo:

$$E = P_r(y = 0|D = 0) = \frac{P_r(y = 0 \cap D = 0)}{P_r(D = 0)}$$

De la misma forma, siendo ésta su definición poblacional, podemos aproximarla con la muestral. Aplicando el axioma anterior obtenemos que:

$$E = \frac{V_-}{V_- + F_+}$$

Conocida como la *fracción de verdaderos negativos*(FVN) o exactitud negativa. Por otro lado, para la construcción de la curva ROC usaremos su complementario:

$$1 - E = 1 - \frac{V_-}{V_- + F_+} = \frac{V_- + F_+}{V_- + F_+} - \frac{V_-}{V_- + F_+} = \frac{F_+}{V_- + F_+}$$

Llamada *fracción de falsos positivos(FFP).*

Observación. Si $E = 1 \Rightarrow F_+ = 0$.

Las medidas de sensibilidad y especificidad corresponden a una *probabilidad a priori*, es decir, dado el estado real de un individuo calculan la probabilidad de que este obtenga un resultado positivo o negativo en la prueba. Puede ser útil en la práctica calcular dichas *probabilidades a posteriori*, es decir, dado el resultado de la prueba de un individuo, cuál será la probabilidad de que presente o no el evento de interés.

Para individuos con resultados positivos en la prueba, dicha probabilidad a posteriori es:

$$P_r(D = 1|y = 1) =$$

$$= \frac{P_r(y = 1|D = 1) \cdot P_r(D = 1)}{P_r(y = 1|D = 1) \cdot P_r(D = 1) + P_r(y = 1|D = 0) \cdot P_r(D = 0)} =$$

$$= \frac{S \cdot p}{S \cdot p + (1 - E) \cdot (1 - p)}$$

Cuya estimación muestral se calcula mediante el *valor predictivo positivo* o *precisión* de la variable de decisión:

$$VPP = \frac{V_+}{V_+ + F_+}$$

Y, para individuos con resultados negativos en la prueba se tiene:

$$P_r(D=0|y=0) =$$

$$= \frac{P_r(y=0|D=0)\cdot P_r(D=0)}{P_r(y=0|D=0)\cdot P_r(D=0) + P_r(y=0|D=1)\cdot P_r(D=1)} =$$

$$= \frac{E\cdot(1-p)}{(1-S)\cdot p + E\cdot(1-p)}$$

Con el *valor predictivo negativo* como su estimación muestral.

$$VPN = \frac{V_-}{V_- + F_-}$$

Cuanto más cercanas a 1 sean éstas probabilidades, mejor capacidad discriminante tendrá nuestra variable de decisión.

Definición. El *valor global de resultados válidos* es la proporción de resultados verdaderos, tanto positivos como negativos:

$$VG = \frac{V_+ + V_-}{V_+ + F_+ + V_- + F_-}$$

Observación. Usando las definiciones muestrales anteriores se obtiene

$$FVP + FFN = P_r(y=1|D=1) + P_r(y=0|D=1) = 1$$

y

$$FFP + FVN = P_r(y=1|D=0) + P_r(y=0|D=0) = 1$$

La siguiente tabla recoge los anteriores resultados obtenidos:

	$y=1$	$y=0$	$Total$
$D=1$	$FVP = \frac{V_+}{V_+ + F_-}$	$FFN = \frac{F_-}{V_+ + F_-}$	1
$D=0$	$FFP = \frac{F_+}{V_- + F_+}$	$FVN = \frac{V_-}{V_- + F_+}$	1

Ya tenemos, por tanto, todos los elementos para representar la curva ROC.

2.2. Definición de la curva ROC

Definición. La *curva ROC poblacional* representa *1-especificidad* frente a la *sensibilidad* para cada posible valor umbral o punto de corte en la escala de resultados de la prueba en estudio. Es decir, $y = f(x)$, donde:

$$ROC(c) = \begin{cases} y = S(c) \\ x = 1 - E(c) \end{cases}$$

Sin embargo, ante la dificultad de obtener datos poblacionales, podemos aproximarla por la *curva ROC muestral*, que representa la *fracción de falsos positivos* en abcisas frente a la *fracción de verdaderos positivos* en ordenadas.

$$ROC_p(c) = \begin{cases} y = FVP(c) \\ x = FFP(c) \end{cases}$$

Puesto que en ambos ejes tenemos probabilidades, la curva ROC, tanto muestral como poblacional, estará contenía en el cuadrado $[0,1] \times [0,1]$. Además, por convenio, se considera que los enfermos tienen valores de x, en general, mayores que los sanos. Por tanto, la curva estará contenida en el triángulo: $\{(x,y)|0 \leq x \leq y \leq 1\}$. Si por la naturaleza de la prueba los resultados estuviesen invertidos (enfermos dan, en media, valores más bajos que sanos) habría que reordenarlos. En el Capítulo 3 se aborda este aspecto con más detalle.

Proposición. *Sean las variables $x_E = (x|D = 1)$ y $x_S = (x|D = 0)$ la variable aleatoria de decisión condicionada al grupo de enfermos y, por otro lado, al grupo de sanos. Sus correspondiente funciones de distribución son: $F_E(x) = P_r(x_E \leq x)$ y $F_S(x) = P_r(x_S \leq x)$ respectivamente. Suponiendo que el valor de x es, por lo general, mayor en individuos con el evento de interés. Se define por tanto, la curva ROC asociada a la variable x como la función:*

$$ROC(t) = 1 - F_E(F_S^{-1}(1-t)) \quad 0 \leq t \leq 1$$

donde t es el complementario de la especificidad y $ROC(t)$ la sensibilidad:

$$\begin{cases} t = 1 - E = 1 - F_S(x_0) \\ q = S = 1 - F_E(x_0) \end{cases}$$

para cada posible x_0 valor de punto de corte.

Demostración. A partir de la definición de la especificidad:

$$E = P\left(y = 0|D = 0\right) \Rightarrow E(u) = F_S(x \leq u) = F_S(u) \Rightarrow$$
$$\Rightarrow (1 - E)(u) = F_S(u > t) = 1 - F_S(u)$$

y, por otro lado, a partir de la definición de sensibilidad:

$$S = P\left(y = 1|D = 1\right) \Rightarrow S(u) = F_E(x > u) = 1 - F_E(u)$$

ahora bien, para cada t la curva ROC representa el par:

$$(1 - E, S) = (1 - F_S(u), 1 - F_E(u)) \Rightarrow$$
$$t = 1 - F_S(u) \Rightarrow u = F_S^{-1}(1-t) \Rightarrow$$
$$ROC(t) = 1 - F_E(F_S^{-1}(1-t))$$

□

Dado que la sensibilidad y la especificidad usadas son unas estimaciones, cabe preguntarse cuánto de cerca estamos de sus correspondientes valores reales. Para ello construiremos sus *intervalos de confianza*. Éstos serán hechos con el método clásico para *proporciones* puesto que, tanto sensibilidad como especificidad, lo son. Por tanto, usaremos como estadístico pivote:

$$Z = \frac{\hat{p_r} - p_r}{SD(p_r)} = \frac{\hat{p_r} - p_r}{\sqrt{\hat{p_r}\hat{q_r}/n}}$$

siendo p_r la proporción a estimar y SD su desviación estándar que, en el caso de la sensibilidad, queda:

$$SD(S) = \sqrt{\frac{\frac{V_+}{(V_+ + F_-)} \cdot \frac{F_-}{(V_+ + F_-)}}{(V_+ + F_-)}} = \sqrt{\frac{V_+ \cdot F_-}{(V_+ + F_-)^3}}$$

y en el caso de la especificidad:

$$SD(E) = \sqrt{\frac{\frac{V_-}{(V_- + F_+)} \cdot \frac{F_+}{(V_- + F_+)}}{(V_- + F_+)}} = \sqrt{\frac{V_- \cdot F_+}{(V_- + F_+)^3}}$$

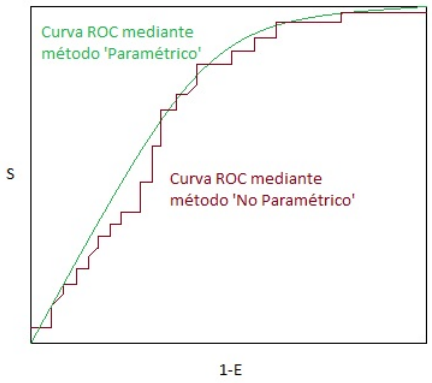
sustituyendo obtenemos que:

$$IC(S) = FVP \pm Z_{\alpha/2} \cdot \sqrt{\frac{V_+ \cdot F_-}{(V_+ + F_-)^3}}$$

$$IC(E) = FVN \pm Z_{\alpha/2} \cdot \sqrt{\frac{V_- \cdot F_+}{(V_- + F_+)^3}}$$

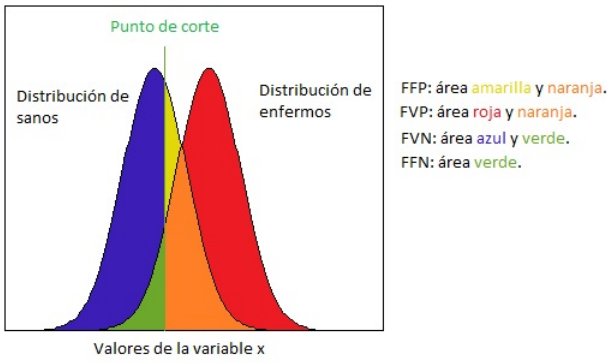
Por último, hay que tener en cuenta que, dado los resultados de la prueba, no siempre vamos a saber qué distribución siguen o bajo qué parámetros. Ésto nos puede llevar a tener curvas escalonada (*método no paramétrico*) o curvas suaves habiendo supuesto la distribución previamente (*método paramétrico*).
En la figura 2.2.1 se aprecia una misma prueba representada mediante método paramétrico y no paramétrico. En principio no sabemos cual se acerca más a la real curva ROC poblacional pero al haber puntos en los que los trazos se separan podemos determinar que nuestro estudio no tendrá los mismos resultados si trabajamos con una o con otra. Por tanto, es importante saber elegir, según cada situación, bajo qué método nos conviene más estimar la curva ROC.

Figura 2.2.1: Ejemplo de curva ROC paramétrica y no paramétrica superpuestas.



Observación. Mientras que la variable x tome valores más altos en enfermos que en sanos, la curva ROC tendrá una *curvatura cóncava* y será *monótona no decreciente*. Este hecho se debe a que si tomamos el valor más bajo posible de punto de corte y avanzamos hasta el mayor, la sensibilidad va a disminuir mientras que la especificidad va a aumentar, y puesto que se representa el par $(1 - E, S)$ nos da una curva por encima o igual que la diagonal $y = x$ del cuadrado $[0, 1] \times [0, 1]$. En la figura 2.0.3 podemos ver cómo el estimador de sensibilidad: FVP, y el del complementario de la especificidad: FFP, disminuyen ambos al mover el punto de corte hacia la derecha.

Figura 2.2.2: Fracciones de falsos y verdaderos positivos y negativos sobre ejemplo de distribuciones de enfermos y sanos.



2.3. Métodos no paramétricos

El método no paramétrico para la construcción de la curva ROC no hace suposición alguna sobre la distribución de los resultados de la prueba en ambos grupos. Debido a ésta característica J.M.Vivo y M.Franco[42] afirman que este método tiene la propiedad de robustez.

Método Empírico

En el método empírico se hace uso de las *funciones de distribución empíricas:* $\hat{F}_{n_E}$ y $\hat{F}_{n_S}$ de los individuos que presentan el evento y los que no, respectivamente, para la construcción de la curva ROC.

Definición. Dada una muestra aleatoria simple $x_1, ..., x_n$ asociada a la variable aleatoria x con función distribución F, se define la *función de distribución empírica* asociada a la muestra como:

$$\hat{F}_n : \mathbb{R} \longrightarrow [0,1]$$

$$x \longrightarrow \hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i(x)$$

donde:

$$\varepsilon_i(x) = \begin{cases} 1 & x_i \leq x \\ 0 & x_i > x \end{cases}$$

Es decir, mide la proporción de observaciones menores que un cierto valor fijado.

En nuestro caso, las expresiones de ambas funciones de distribución empíricas serán:

$$\hat{F}_{n_E}(x) = \frac{1}{n_E}\sum_{i=1}^{n_E} \varepsilon_i(x)$$

$$\hat{F}_{n_S}(x) = \frac{1}{n_S}\sum_{i=1}^{n_S} \varepsilon_i(x)$$

siendo $\{x_1, ..., x_{n_E}\}$ los individuos que presentan el evento en la muestra y $\{x_1, ..., x_{n_S}\}$ los que no.

Definición. Se define la curva ROC empírica como la curva construida uniendo los $\left(1 - \hat{F}_{n_S}(x), 1 - \hat{F}_{n_E}(x)\right)$ consecutivos.

$$\widehat{ROC}_n(t) = 1 - \hat{F}_{n_E}(1 - \hat{F}_{n_S}(x))$$

siendo n el tamaño de la muestra.

Este método nos da una curva ROC escalonada donde puede haber trazos verticales, horizontales o diagonales, cada uno indica qué ha ocurrido al variar el punto de corte:

- El trazo horizontal indica que un caso ha pasado a ser F_+
- El trazo vertical indica que un caso ha pasado a ser V_+
- El trazo diagonal indica que ha habido un empate entre dos individuos de distintos grupos, es decir, que un enfermo y un sano han pasado a ser ambos F_+ ó V_+

El método empírico es el recomendado para pruebas con resultados continuos, en especial si el redondeo (la discretización de la escala) es escasa ya que nos dará pocos empates (trazos diagonales) en la curva ROC. Además, los escalones tienden a suavizarse conforme aumenta el tamaño de la muestra, convergiendo así a la real curva ROC poblacional.

Por otro lado, también se puede aplicar a pruebas con resultados de tipo *categóricos*. Nótese que en este caso vamos a tener empates siempre que haya más individuos que categorías y la curva serán los pares $(1-E, S)$ calculados para cada categoría como punto de corte.

Método de la función Kernel

Una forma de evitar obtener una curva escalonada, además de con procedimientos paramétricos, es usando la *función Kernel* para estimar las distribuciones de la variable de decisión en ambos grupos. Este procedimiento también recibe el nombre de *estimación suavizada de la curva ROC* cuyo primer estudio fue desarrollado por Zou et al.[30] y más tarde por Lloyd[31].

Definición. Sean $\{x_{1_E}, ..., x_{n_E}\}$ los individuos de la muestra que presentan el evento, y sean $\{x_{1_S}, ..., x_{m_S}\}$ los que no. Se definen las *funciones de densidad kernel* estimadas como:

$$\tilde{f}_E(x) = \frac{1}{n_E \cdot h_1} \sum_{i=1}^{n_E} K_1\left(\frac{x - x_i}{h_1}\right)$$

$$\tilde{f}_S(x) = \frac{1}{m_S \cdot h_2} \sum_{i=1}^{n_S} K_2\left(\frac{x - x_i}{h_2}\right)$$

donde h_1 y h_2 es una secuencia de números positivos llamada *ancho de banda* o *bandwidths* que determina cuánto de suavizada queda la curva.
La elección de un ancho de banda adecuado ha sido estudiada por Lloyd y Yong[37], Zhou y Harezlak[38], Hall y Hyndmann[39] y Jokiel-Rokita y Pulit [40]. En concreto, Lloyd y Yong[37] propusieron métodos empíricos para estimarlos, calculando así por separado ambas distribuciones. Sin embargo, Hall y Hyndman[39] desarrollaron una técnica para estimar el ancho de banda que permitía calcular las distribuciones en conjunto, es decir, usando las condicionadas a presentar o no el evento de interés. Las funciones Kernel son continuas, simétricas y positivas, además cumplen:

- $\int_{\mathbb{R}} K_i(x)\delta x = 1$
- $\int_{\mathbb{R}} x \cdot K_i(x)\delta x = 0$
- $\int_{\mathbb{R}} x^2 \cdot K_i(x)\delta x > 0$

para $i = 1, 2$.
Las expresiones de las funciones de distribución correspondientes a las anteriores funciones de densidad son:

$$\tilde{F}_E(x) = \frac{1}{n_E}\sum_{i=1}^{n_E}\int_{-\infty}^{x}\frac{1}{h_1}K_1\left(\frac{u - x_{i_E}}{h_1}\right)\delta u$$

$$\tilde{F}_S(x) = \frac{1}{m_S}\sum_{i=1}^{m_S}\int_{-\infty}^{x}\frac{1}{h_2}K_2\left(\frac{u - x_{i_S}}{h_2}\right)\delta u$$

Zou et. al[48] proponen hacer una estimación *gaussiana-kernel*, que es la que sigue:

$$\tilde{F}_E(x) = \frac{1}{n_E}\sum_{i=1}^{n_E}\Phi\left(\frac{x - x_{i_E}}{h_1}\right)$$

$$\tilde{F}_S(x) = \frac{1}{m_S}\sum_{i=1}^{m_S}\Phi\left(\frac{x - x_{i_S}}{h_2}\right)$$

siendo Φ la función de densidad de una $N(0, 1)$. Finalmente, bajo éstas estimaciones, la expresión de la curva ROC por el método Kernel es:

$$\widetilde{ROC}(t) = 1 - \tilde{F}_E\left(\tilde{F}_S^{-1}(1 - t)\right)$$

2.4. Métodos paramétricos

El método paramétrico trata de averiguar la distribución de la variable de decisión o variable respuesta de la prueba. Como ya hemos dicho anteriormente, dicha distribución no tiene porqué ser la misma en el grupo de sanos y enfermos. De hecho, nuestro caso óptimo es en el que las gráficas de las distribuciones no se solapen, es decir, que estén lo más alejadas posible. Esto se traduciría en una alta capacidad discriminante. Sin embargo, si estuviesen muy solapadas o, directamente fuesen iguales, la capacidad discriminante prácticamente sería nula.

Es recomendable aplicar transformaciones monótonas sobre los datos de la muestra con idea de que se ajuste mejor a la distribución deseada, esto no añade error a nuestros resultados pues la curva ROC es invariante bajo este tipo de transformaciones ya que lo que representa son porcentajes[7].

Existen varios métodos que ajustan la variable a distribuciones como la *logística*

o la *exponencial negativa* éstos han sido propuestos por Zweig y Campbell[32], pero la distribución más usada es el de la *normal*, que supone la *normalidad* de la variable en ambos grupos.

Los primeros en usar este modelo fueron Swets y Green[1], Dorfman y Alf[33] y más tarde Turnbull[34]. Siendo Hanley[8]el primero en publicar un artículo exponiendo las razones por las cuales es tan *robusto* este modelo, pues afirmaba que, aun teniendo una cierta cantidad de observaciones que no siguen una distribución normal, ofrece buenos resultados. Además, según afirman Heritier[36] y Farcomeni y Ventura[35], los modelos teóricos son sólo una aproximación de la realidad y son necesarios todos aquellos procedimientos estadísticos que salven esta diferencia[24].

Proposición. *La curva ROC estimada mediante el modelo binormal se determina con dos parámetros, siendo éstos:*

$$\hat{a} = \frac{\hat{\mu}_E - \hat{\mu}_S}{\hat{\sigma}_E}$$

$$\hat{b} = \frac{\hat{\sigma}_S}{\hat{\sigma}_E}$$

Con $\hat{\mu}_E$, $\hat{\sigma}_E$, $\hat{\mu}_S$, $\hat{\sigma}_S$ las medias y desviaciones típicas estimadas de enfermos y sanos respectivamente. Quedando la curva determinada por la expresión:

$$ROC(t) = 1 - \Phi(\hat{a} + \hat{b} \cdot \Phi^{-1}(1-t)) \quad 0 \leq t \leq 1$$

Donde Φ representa la función de distribución de la normal estándar.

Una vez elegido el modelo, el siguiente paso es estimar los parámetros de dicha distribución y la manera más usada de hacerlo es mediante el *método de máxima verosimilitud*. Metz et al.[23] publicó un artículo con las estimaciones de máxima verosimilitud de varias distribuciones continuas aplicadas a las curvas ROC. Una vez que tenemos la distribución junto con sus parámetros determinadas lo que le sigue es la representación de la curva ROC. Ésta será suave, no escalonada, al contrario de las obtenidas con el modelo empírico, pero nos arriesgamos a cometer mucho *error* si la elección de la distribución no ha sido la acertada. Para ello se recomienda efectuar un *contraste de hipótesis de bondad de ajuste*, donde la *hipótesis nula* es que la variable sigue la distribución elegida y la *hipótesis alternativa* es que difiere de ella.

Por último, también es recomendable que, si la variable no se ajusta a ninguna distribución, hacerle transformaciones con el objetivo de que el ajuste sea más acertado.

2.5. Método semiparamétrico

Este método nace ante la popularidad que tomó el *modelo binormal* tras, que el investigador J.A. Hanley[8], advirtiese de sus múltiples usos en referencia al

ajuste de variables.
Lo primero que hace este método es agrupar los datos en categorías ordenadas y luego aplicarle el modelo binormal. El nombre de semiparamétrico le viene porque asume la normalidad de la variable pero no especifica qué transformación hace que la variable siga realmente una normal. Este último paso se respalda, como hemos dicho en la sección anterior, en que la curva ROC es invariante ante transformaciones monótonas de la variable de decisión[7]. En efecto, sea $ROC(t) = 1 - F_E(F_S^{-1}(1-t))$ la curva ROC asociada a una variable de decisión x con funciones de distribución asociadas $F_E(\cdot)$ y $F_S(\cdot)$ en enfermos y sanos respectivamente, sea T una transformación monótona y sea $w = T(x)$, que por ser monótona existe inversa:

$$F_{w,E}(t) = F_E(T^{-1}(t))$$

$$F_{w,S}(t) = F_S(T^{-1}(t))$$

entonces la curva ROC asociada a w es:

$$ROC_w(t) = 1 - F_{w,E}(F_{w,S}^{-1}(1-t)) =$$

$$= 1 - F_{w,E}(T(F_S^{-1}(1-t))) =$$
$$= 1 - F_E(T^{-1}(T(F_S^{-1}(1-t)))) =$$
$$= 1 - F_E(F_S^{-1}(1-t)) = ROC(t)$$

Como la transformación de la variable de decisión sigue una distribución normal, podemos sustituir quedándonos la siguiente expresión:

$$ROC(t) = 1 - F_E(F_S^{-1}(1-t))$$

$$ROC(t) = 1 - F_E T T^{-1} F_S^{-1}(1-t) =$$
$$= 1 - \Phi\left(\frac{\mu_S + \sigma_S \Phi^{-1}(1-t) - \mu_E}{\sigma_E}\right) =$$

siendo μ_S y μ_E las medias de $T(x_S)$ y $T(x_E)$ respectivamente y las desviaciones típicas σ_S y σ_E.

$$= 1 - \Phi(\hat{a} + \hat{b} \cdot \Phi^{-1}(1-t))$$

Es decir, la expresión para la curva ROC estimada por el método binormal. Cai y Pepe[41] estudian las propiedades de esta estimación y realizan estudios de simulación utilizando tanto métodos paramétricos como semiparamétricos.

Comparación de curvas estimadas por el modelo binormal

Dadas dos curvas ROC es recomendable someter a contraste de hipótesis la igualdad entre ellas, es decir, dadas dos variables de descisión x_1 y x_2, ¿tienen la misma capacidad discriminante?, ¿a cada valor de $1 - E$ le corresponde el mismo valor de S en ambas curvas? o, análogamente con sus estimadores muestrales, ¿a cada valor de la FFP le corresponde el mismo valor de la FVP?

Estas preguntas podrían responderse con un análisis gráfico, la camparación del área que encierren bajo la curva o algún otro cuantificador de exactitud, que veremos en el capítulo 3, pero Vivo y Franco[42] sostienen que es necesario hacer el contraste de hipótesis para aceptar o no la igualdad de clasificadores.

En general, para cualquier modelo de estimación la hipótesis nula sería si para cada FFP se tiene el mismo valor de FVP. En el caso concreto del modelo binormal, puesto que la curva está determinada por dos paramétros, haremos de nuestra hipótesis nula la igualdad de éstos parámetros, es decir, sean las curvas $ROC_1(t)$ y $ROC_2(t)$ cuyos parámetros son $(\hat{a}_1, \hat{b}_1)$ y $(\hat{a}_2, \hat{b}_2)$ respectivamente.

$$\begin{cases} H_0 : \hat{a}_1 = \hat{a}_2 \wedge \hat{b}_1 = \hat{b}_2 \\ H_1 : \hat{a}_1 \neq \hat{a}_2 \vee \hat{b}_1 \neq \hat{b}_2 \end{cases}$$

El estadístico que resuelve este contraste es:

$$\chi^2 = \frac{\hat{a}_{12} \cdot \hat{V}(\hat{b}_{12}) + \hat{b}_{12}^2 \cdot \hat{V}(\hat{a}_{12}) - 2 \cdot \hat{a}_{12} \cdot \hat{b}_{12} \cdot cov(\hat{a}_{12}, \hat{b}_{12})}{\hat{V}(\hat{a}_{12}) \cdot \hat{V}(\hat{b}_{12}) - cov(\hat{a}_{12}, \hat{b}_{12})}$$

propuesto por Metz y Kronman[43], donde $\hat{a}_{12} = \hat{a}_1 - \hat{a}_2$ y $\hat{b}_{12} = \hat{b}_1 - \hat{b}_2$ y que tiende asintóticamente a una distribución chi-cuadrado de 2 grados de libertad bajo hipótesis nula. Por tanto, para un nivel de significación α se rechaza la igualdad de pruebas si $\chi^2 > \chi^2_{2,\alpha}$.

Capítulo 3

Medidas de exactitud para un clasificador

Una vez dibujada la curva ROC necesitaremos una medida con la cual interpretar si la variable de estudio discrimina bien a nuestra muestra, o si, dada otra variable, clasifica mejor o peor.

3.1. Precisión o exactitud de un clasificador

Definición. Se define la exactitud o acuracidad, *accuracy(AC),* de una variable de decisión como la probabilidad de discriminar correctamente.
Por el teorema de la probabilidad total se puede obtener la relación de este concepto con la sensibilidad y especificidad:

$$AC = P_r(y = 1|D = 1) \cdot P_r(D = 1) + P_r(y = 0|D = 0) \cdot P_r(D = 0) \Rightarrow$$

$$AC = S \cdot P_r(D = 1) + E \cdot P_r(D = 0) \Rightarrow$$

$$AC = S \cdot prevalencia + E \cdot (1 - prevalencia)$$

Es, por tanto, una suma ponderada de sensibilidad y especificidad con pesos de prevalencia y su complementario. Dada una muestra, un estimador basado en las frecuencias observadas es:

$$\hat{AC} = \frac{V_+ + V_-}{V_+ + V_- + F_+ + F_-} = \frac{resultados\, acertados}{total\, de\, la\, muestra}$$

Cuanto más se aproxime la cantidad de resultados acertados al total de individuos de la muestra, más cercano a 1 será la exactitud y tendremos, por tanto,

una prueba con una alta capacidad discriminante. Esto significaría que apenas tendríamos F_+y F_-, es decir, resultados erróneos.

$$Si\,(F_+ + F_-) \to 0 \Rightarrow \hat{AC} \to 1$$

Por otro lado, de la fórmula anterior, se deduce un estimador para la prevalencia llamado *predominancia(PD)* que mide la cantidad de individuos de la muestra que presenta el evento de estudio:

$$PD = \frac{\hat{AC} - FVN}{FVP - FVN}$$

donde FVN es la fracción de verdaderos negativos, el estimador muestral de la especificidad, y FVP la fracción de verdaderos positivos, el estimador de la sensibilidad.

Esta medida responde a la pregunta de qué porcentaje es el acertado de entre todos los positivos que ha clasificado la prueba. Sin embargo, es insuficiente para conocer la bondad de la misma puesto que ajustar un clasificador es cuestión de equilibrar en términos de aquello que sea más importante o relevante para el objeto de investigación, a saber, la precisión, la sensibilidad o la especificidad. No siempre es posible optimizar conjuntamente todas estas medidas y se ha de priorizar alguna de ellas.

3.2. Índice de Youden

El índice de exactitud anterior mide la proporción de pacientes correctamente clasificados sin hacer diferencia entre positivos y negativos. Otra medida propuesta es el denominado *Índice de Youden* [29]. Este refleja la diferencia entre la tasa de verdaderos positivos y la de falsos positivos. Un buen test debe tener alta esta diferencia. O de forma equivalente, debe tener una alta especificadad y una alta especificidad.

Definición. Se define el *índice de Youden* como la diferencia entre las respuestas positivas correctas y las respuestas positivas incorrectas, es decir, los positivos de la prueba en ausencia del evento de interés:

$$\gamma = Pr(y = 1|D = 1) + Pr(y = 1|D = 0)$$

o equivalentemente,

$$\gamma = S + E - 1$$

pudiéndose escribir su estimador como:

$$\hat{\gamma} = FVP - FFP$$

Este índice toma valores en el intervalo $[0, 1]$ siendo los valores cercanos a 0 los correspondientes a una prueba con poca capacidad discriminatoria y los cercanos a 1 a una prueba perfecta pues significaría que $S = 1$ y $E = 1$.

3.3. Tasas de verosimilitud

Definición. Se define la *tasa o razón de verosimilitud(LR)* como el cociente de la probabilidad de respuesta positiva, o negativa, bajo presencia del evento entre la probabilidad de respuesta positiva, o negativa, bajo ausencia del evento de interés. Así, las tasas de verosimilitud positiva y negativa se definen como:

- *Tasa de verosimilitud positiva o likehood ratio positivo(LRP)*, en caso de que la respuesta de la prueba sea positiva:

$$LRP = \frac{P_r(y = 1|D = 1)}{P_r(y = 1|D = 0)}$$

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

- *Tasa de verosimilitud negativa o likehood ratio negativo(LRN)*, en caso contrario, que la prueba dé respuesta negativa:

$$LRN = \frac{P_r(y = 0|D = 1)}{P_r(y = 0|D = 0)}$$

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

Los valores de LRP y LRN recorren el intervalo $[0, +\infty)$. Un valor mayor que 1 significa que hay más respuestas positivas en individuos que presenten el evento que en los que no, y viceversa para un valor menor que 1. El cuadro 3.1 recoje una clasificación de la capacidad discriminante de la prueba en función del valor de LR, propuesta por Jaeschke *et al.*[49]. En el ámbito de la epidemiología, la tasa LRP proporciona una medida de cuántas veces es más probable que la prueba sea positiva en los enfermos que en los sanos. Análogamente, LRN proporciona una medida de cuántas veces es más probable que la prueba sea negativa en enfermos que en sanos.
A partir de los datos muestrales se proponen los siguiente estimadores de las tasas de verosimilitud:

- Estimador de la tasa de veroimilitud positiva:

$$\widehat{LRP} = \frac{FVP}{FFP}$$

- Estimador de la tasa de verosimilitud negativa:

$$\widehat{LRN} = \frac{FFN}{FVN}$$

Cuadro 3.1: Categorización de la tasa de verosimilitud[16]

$< 0{,}1$	$0{,}1 - 0{,}2$	$0{,}2 - 0{,}5$	$0{,}5 - 2$	$2 - 5$	$5 - 10$	> 10
Excelente	Muy bueno	Bueno	Justo	Bueno	Muy bueno	Excelente

Al ser la razón de verosimilitud definida con sensibilidad y especificidad está sujeta al error que puedan proporcionar la FVP y la FVN. Podemos cuantificar dicho error mediante el error estándar, cuya expresión para cada tipo de tasa viene dada por:

$$SD(\widehat{LRP}) = \sqrt{\frac{FFN}{V_+} + \frac{FVN}{F_+}}$$

$$SD(\widehat{LRN}) = \sqrt{\frac{FVP}{F_-} + \frac{FFP}{V_-}}$$

Siendo sus intervalos de confianza:

$$IC(LRP) = \frac{FVP}{FFP} \pm exp(Z_{1-\alpha/2} \cdot SD(\widehat{LRP}))$$

y

$$IC(LRN) = \frac{FFN}{FVN} \pm exp(Z_{1-\alpha/2} \cdot SD(\widehat{LRP}))$$

3.4. Odds ratio

De forma semejante al concepto de odds ratio definido como medida de asociación entre dos variables dicotómicas, se puede definir como una medida de la capacidad de precisión de un test. En este caso vamos a cuantificar la probabilidad de respuesta positiva en la prueba, es decir $y = 1$, tanto en individuos con el evento $D = 1$, como en individuos sin él $D = 0$.

Definición. Se define el *odds* de un suceso A como el siguiente cociente de probabilidades:

$$Odds(A) = \frac{P_r(A)}{1 - P_r(A)} = \frac{P_r(A)}{P_r(\bar{A})}$$

Así, la ventaja u oportunidad de un suceso puede interpretarse como sigue:

- $Odds(A) = \frac{P_r(A)}{P_r(\bar{A})} > 1 \Rightarrow$La probabilidad de ocurrencia del evento A es mayor que la no ocurrencia.
- $Odds(A) = \frac{P_r(A)}{P_r(\bar{A})} < 1 \Rightarrow$La probabilidad de ocurrencia de A es menor que la no ocurrencia.
- $Odds(A) = \frac{P_r(A)}{P_r(\bar{A})} = 1 \Rightarrow$Hay igual probabilidad de ocurrencia del evento A que de su complementario.

Siguiendo esta expresión, podemos hacer dos definiciones, una cuando haya presencia del evento y otra cuando no. De manera que nos quedaría:

- Ocurrencia de respuesta positiva en presencia del evento:

$$Odds_{presencia} = \frac{P_r(y = 1|D = 1)}{P_r(y = 0|D = 1)}$$

$$Odds_{presencia} = \frac{S}{1 - S}$$

- Ocurrencia de respuesta positiva en ausencia del evento:

$$Odds_{ausencia} = \frac{P_r(y = 1|D = 0)}{P_r(y = 0|D = 0)}$$

$$Odds_{ausencia} = \frac{1 - E}{E}$$

La capacidad diagnóstica de una prueba será mejor cuanto mayor sea $Odds_{presencia}$ en comparación con $Odds_{ausencia}$, por ello, una medida de la precisión del test es la razón entre ambas.

Definición. Se define el *odd ratio* como el siguiente cociente de *odds*:

$$Odds\, ratio = \frac{Odds_{presencia}}{Odds_{ausencia}}$$

$$Odds\, ratio = \frac{S \cdot E}{(1 - S) \cdot (1 - E)}$$

Siendo por tanto su estimador:

$$\widehat{OD} = \frac{V_+ \cdot V_-}{F_- \cdot F_+}$$

Según sus valores indican:

- $Odds\, ratio > 1$ ⇒Mayor ocurrencia de respuesta positiva cuando el evento de interés está presente.
- $Odds\, ratio < 1$ ⇒Menor ocurrencia de respuesta positiva cuando el evento está presente.
- $Odds\, ratio = 1$ ⇒Igual ocurrencia en ambos casos.

3.5. Índice de discriminación

Si la variable del test x es cuantitativa, la capacidad de un test basado en dicha variable se puede determinar por la cantidad de *solapamiento* entre las distribuciones de probabilidad de x, en caso de evento $D = 1$, y en caso de no evento, $D = 0$. En concreto, cuanto más solapamiento, menor capacidad de discriminación y viceversa.

En caso de bi-normalidad, es decir:

$$\begin{cases} x|D = 0 \sim N(\mu_1, \sigma) \\ x|D = 1 \sim N(\mu_2, \sigma) \end{cases}$$

Cuanto mayor separación haya entre las medias, menor solapamiento y, por tanto, más capacidad de discriminación. Por ello, se puede utilizar como una medida de capacidad de discriminación, la aseparación en términos relativos de la dispersión de ambas distribuciones.

Definición. El *índice de discriminación* es una medida para cuantificar cómo de separadas, gráficamente, están las funciones de densidad del grupo de individuos que presenta el evento de interés frente al que no lo presenta. Su expresión viene dada por:

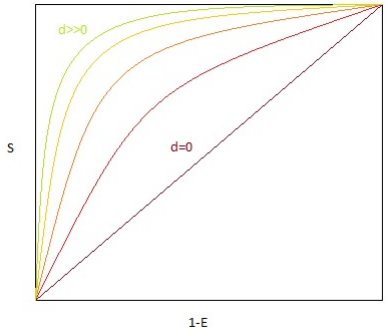
$$\delta = \frac{separación}{dispersión} = \frac{|\mu_E - \mu_S|}{\sigma}$$

Siendo μ_E y μ_S las medias de nuestra variable de descisión: 'x' en el grupo que presenta el evento y en el grupo que no lo presenta respectivamente. Por otro lado, σ representa el error estandar de la variable de decisión en la muestra total. Es decir:

$$\begin{cases} \mu_E = \overline{x_E} \\ \mu_S = \overline{x_S} \\ \sigma = SD(x) \end{cases}$$

Este índice mide la proporción de pacientes correctamente diagnosticados, pero trata por igual a positivos y negativos, verdaderos o falsos. Los valores que toma ésta medida recorre el intervalo $[0, +\infty)$, siendo 0 el correspondiente a una capacidad discriminante nula y, por tanto, a una curva ROC solapada con el segmento de la recta $y = x$ contenida en el cuadrado $[0, 1] \times [0, 1]$.

Figura 3.5.1: Ejemplo de curvas ROC en función de su índice de discriminalidad.



3.6. Área bajo la curva

El *área bajo la curva* es el estadístico por excelencia para medir la capacidad discriminante de la prueba. También para comparar pruebas entre sí y determinar cual és la más eficaz.

Definición. Sea $ROC(t)$ la función asociada a la curva ROC. Se define el *área bajo la curva bajo la curva ROC (area under curve, AUC)* como:

$$AUC = \int_0^1 ROC(t)\delta t$$

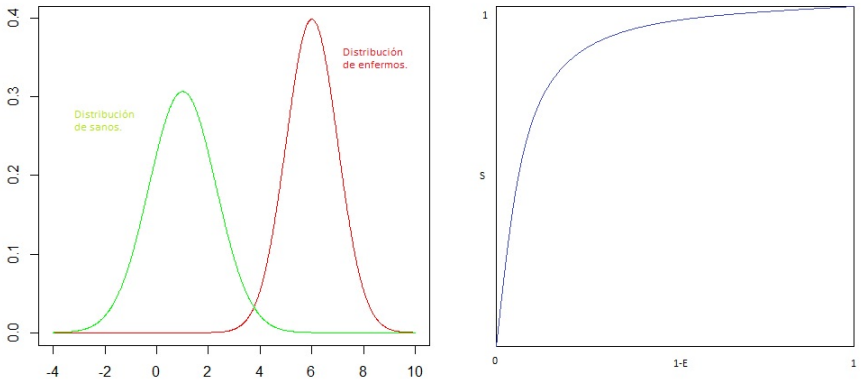
Su rango de valores va desde $0,5$, siendo este valor el correspondiente a una prueba sin capacidad discriminante, hasta 1, que es cuando los dos grupos están perfectamente diferenciados por la prueba. Por tanto, podemos decir que cuanto mayor sea el AUC mejor será la prueba.

Baja exactitud:$[0'5, 0'7)$
Útiles para algunos propósitos:$[0'7, 0'9)$
Exactitud alta:$[0'9, 1]$

Interpretación de Swets[9] para valores del AUC.

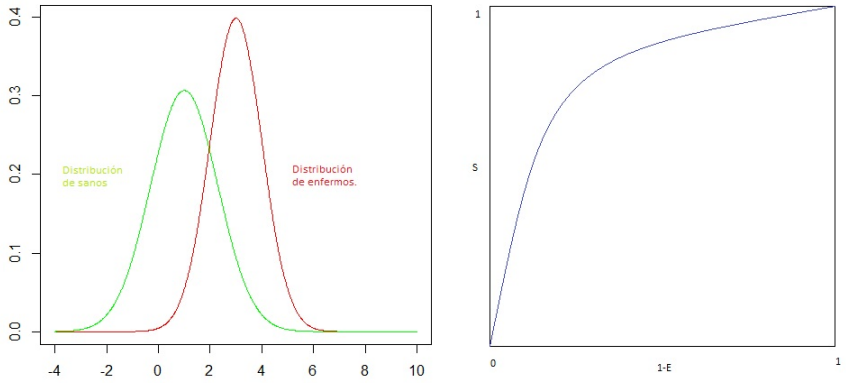
El criterio de Swets[9] dicta que un área por debajo de $0,7$tiene una baja capacidad discriminante y que una por debajo de $0,9$ puede ser útil para algunos propósitos, catalogando a las mayores de $0,9$con alta exactitud.

Figura 3.6.1: Ejemplo de densidades y curva ROC de dos poblaciones bien discriminadas.



En la figura 3.6.1 se representa el caso óptimo en el que las distribuciones de sanos y enfermos se encuentran bien diferenciadas, lo que proporciona una curva ROC muy próxima al punto $(0, 1)$ y por tanto un área casi máxima. La cercanía al $(0, 1)$ es deseable pues significa que para ciertos puntos de corte la prueba obtiene una alta sensibilidad y especificidad, lo que se traduce en bajos resultados erróneos. Sin embargo, en la figura 3.6.2 se aprecia cierto solapamiento, siendo este el caso más frecuente pues, aunque el caso anterior es preferible no es usual encontrarnos con una variable de decisión que apenas ofrezca confusión.

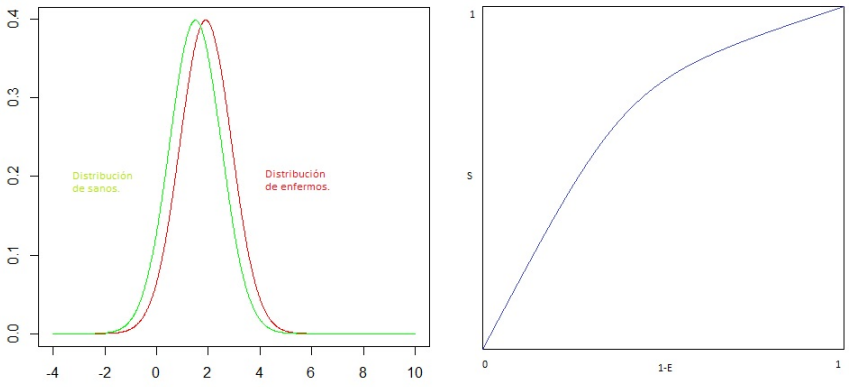
Figura 3.6.2: Ejemplo de densidades y curva ROC de dos poblaciones con cierto solapamiento[6].



Por último, el peor caso que se puede presentar es cuando el solapamiento es

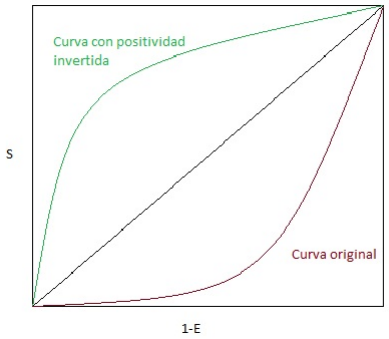
total. Esto quiere decir que la prueba no es útil para detectar la enfermedad o el evento de interés. A la curva ROC le ocurrirá que bajo todo punto de corte $1 - E = S$ y por tanto obtendremos el segmento de la recta $y = x$ contenido en el cuadrado $[0, 1] \times [0, 1]$. Ésto hará que el AUC sea muy cercano a $0,5$ como se ve en la figura 3.6.3.

Figura 3.6.3: Ejemplo de densidad y curva ROC de dos poblaciones completamente solapadas.



Una interpretación del AUC, ofrecida por Hanley y McNeil[44], es la de entenderla como la probabilidad de clasificar correctamente un par de individuos (uno sano y otro enfermo) escogidos al azar. Podríamos traducir esto en que la prueba dé un resultado más anormal en el enfermo que en el sano, es decir, $AUC = P_r(x_S < x_E)$, donde x_S es el valor de la variable de decisión en el grupo de sanos y x_E en el de enfermos.

Figura 3.6.4: Ejemplo de curva ROC simétrica respecto a $y = x$



Observación. Por convenio, se considera que el valor de la variable decisión es mayor en enfermos que en sanos. Si no fuese así, la curva ROC saldría simétrica respecto a la diagonal $y = x$. Para corregir esto basta con invertir la positividad de la prueba, es decir, considerar positivo el evento que se estaba considerando como negativo. Como se puede ver en la figura 3.5.4, este cambio hace que la curva ROC cambie en simetría respecto a la bisectriz del primer cuadrante.

3.7. Cálculo del área bajo la curva

Basándonos en la definición dada de área bajo la curva, una primera forma de calcularla sería representando la curva ROC, o una estimación de ella, y obtener el porcentaje de área del cuadrado $[0,1] \times [0,1]$ que encierra bajo ella. Veremos a continuación que no siempre será necesario tener la curva para obtener este área, sino, que a partir de los datos podremos estimarla directamente.

Método no paramétrico

Regla Trapezoidal

Si usamos el modelo empírico para la construcción de la curva ROC podremos calcular el área como suma de áreas de trapecios puesto que tendremos una curva en forma de escalera.
Su fórmula viene dada por:

$$AUC = \sum_{t=1}^{T} \frac{1}{2}(FFP_t - FFP_{t-1}) \cdot (FVP_t + FVP_{t-1})$$

Siendo (FFP_t, FVP_t) las fracciones de falsos positivos y verdaderos positivos calculadas para cada $t = 1, ..., T$ puntos de corte.

Estadístico suma de rangos de Wilcoxon

Bajo métodos no paramétricos, el AUC también puede estimarse con el *estadístico* w de Wilconxon que, además coincide, con la *suma de rangos* (w) obtenida de la prueba de comparación de medias de Wilconxon. La razón de esta igualdad reside en que el estadístico w es usado para contrastar hipótesis del tipo: $H_0 : P_r(x > y) = 1/2$, que en nuestro caso sería $H_0 : P_r(x_S > x_E) = 1/2$, es decir, que el área fuese $1/2$ y, por tanto, la prueba no distinga entre un grupo y otro. Este resultado fue demostrado por Bamber[10].

Además, a partir de esta relación con w Hanley y McNeil[44] obtuvieron la expresión del error estándar del AUC estimada mediante este estadístico:

$$SD(AUC) = \sqrt{\frac{AUC(1 - AUC) + (n_E - 1)(Q_1 - AUC^2) + (n_S - 1)(Q_2 - AUC^2)}{n_E \cdot n_S}}$$

Siendo:

- Q_1 la probabilidad de que el resultado de la prueba, aplicada a dos individuos del grupo de presencia del evento (enfermo), sea mayor que el resultado de la prueba aplicada a un individuo del grupo de ausencia del evento (sano).

$$Q_1 = P_r(y_{E_i} > y_{S_j},\, y_{E_k} > y_{S_j})$$

para cualquier $\begin{cases} i,k = 1,...,n_E \\ j = 1,...,n_S \end{cases}$

- Q_2 la probabilidad de que el resultado de la prueba, aplicada a dos individuos del grupo de ausencia del evento (sano), sea mayor que el resultado de la prueba aplicada a un individuo del grupo de presencia (enfermo).

$$Q_2 = P_r(y_{S_i} > y_{E_j},\, y_{S_k} > y_{E_j})$$

para cualquier $\begin{cases} i,k = 1,...,n_S \\ j = 1,...,n_E \end{cases}$

Observación. Cuando el número de empates es elevado, como suele ocurrir con los datos de clasificación, el estadístico w deja de ser insesgado para el área. En tal caso sería mejor usar un método paramétrico o semiparamétrico.

Método de la función Kernel

Para las curvas ROC no paramétricas pero suavizadas mediante la función Kernel podemos calcular el área recurriendo a las definiciones:

$$AUC = \int_0^1 ROC(t)\delta t$$

y

$$\widetilde{ROC}(t) = 1 - \tilde{F}_E\left(\tilde{F}_S^{-1}(1-t)\right)$$

de modo que sustituyendo obtenemos:

$$AUC = \int_0^1 \widetilde{ROC}(t)\delta t = \int_0^1 1 - \tilde{F}_E\left(\tilde{F}_S^{-1}(1-t)\right)\delta t$$

Métodos paramétricos y semiparamétricos

Partiendo de que hemos supuesto una distribución y ajustado unos parámetros, cualquier área que calculemos mediante éstos métodos va a estar sujeta a un error. La forma de cuantificar este error es calculando un *intervalo de confianza* que se vea reducido si se ampliase la muestra. Para ello tenemos dos caminos:

- Fijar un punto de corte y calcular un intervalo de confianza para la sensibilidad y la especificidad.
- Fijar la especificidad en un valor, por ejemplo 80 %, y calcular el intervalo de confianza para la sensibilidad o viceversa.

Una vez asegurada la distribución y estimadas $F_E(x)$ y $F_S(x)$ el cálculo numérico del área será, de nuevo, sustituyendo en la definición:

$$AUC = \int_0^1 ROC(t)\delta t = \int_0^1 1 - F_E\left(F_S^{-1}(1-t)\right)\delta t$$

Por otro lado, para el caso concreto del modelo binormal, puesto que la curva queda determinada por los parámetros a y b (véase sección 2,4) podemos obtener la expresión del área en función de ambos.

$$\widehat{AUC} = \int_0^1 \widehat{ROC}(t)\delta t = \int_0^1 1 - \Phi(\hat{a} + \hat{b}\cdot\Phi^{-1}(1-t))\delta t =$$

$$= \Phi\left(\frac{\hat{a}}{\sqrt{1+b^2}}\right) = \Phi\left(\frac{\hat{\mu}_E - \hat{\mu}_S}{\sqrt{\sigma_E^2 + \sigma_S^2}}\right)$$

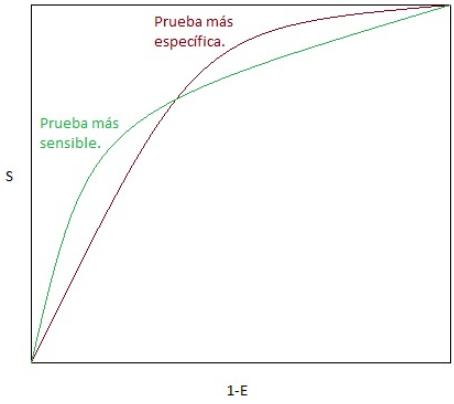
Siendo Φ la función de densidad de una distribución normal estándar.

3.8. *AUC* parcial

Dos curvas ROC iguales tienen el mismo área, sin embargo el viceverso no es cierto. En la figura 3.8.1 ambas curvas tienen igual área pero la verde proporciona una prueba más sensible, es decir, buena para propósitos preventivos a costa de tener una proporción alta de falsos positivos. Por otro lado, la curva roja corresponde a una prueba más específica, es decir, nos deja una proporción alta de falsos negativos. Ésta característica sería deseable para, por ejemplo, tamizaje de enfermedades.

La razón de que una sea más sensible y otra más especifica viene sustentado por la regla de elección del punto de corte. Ésta dice que el óptimo par $(1-E, S)$, en el sentido de minimizar los resultados falsos, es el más cercano al punto $(0, 1)$. En el Capítulo 4 veremos ésta y otras relgas de elección del punto de corte.

Figura 3.8.1: Ejemplo de curvas ROC distintas pero con igual AUC.



Una forma de detectar este caso, a parte de representándola gráficamente, es calculando el área que encierra la curva entre dos verticales dadas, es decir, en una porción del eje horizontal.

Definición. Se define el *área parcial bajo la curva ROC* como el área de la región delimitada entre la curva, el eje horizontal y dos abcisas dadas: $x = a$ y $x = b$, con $a < b$. Siendo su expresión:

$$AUC_{(a,b)} = \int_a^b ROC(t)\delta t$$

La estimación de este valor para curvas ROC suavizadas, bien por el método de Kernel o bien por métodos paramétricos o semiparamétricos es la que sigue, siendo $\widehat{ROC}(t)$ la estimación de la misma:

$$\widehat{AUC}_{(a,b)} = \int_a^b \widehat{ROC}(t)\delta t$$

Para las curvas ROC escalonadas, estimadas mediante el método trapezoidal, la estimación del área parcial es:

$$\widehat{AUC}_{(a,b)} = b_1 + b_2 + \sum_{t=A}^{B} \frac{1}{2}(FFP_t - FFP_{t-1}) \cdot (FVP_t + FVP_{t-1})$$

Siendo:

- $A = min(i - a)/i \geq a$ con $i = 1, ..., T$, donde hasta T son los posibles valores de $1 - E$ para cada punto de corte.
- $B = min(b - j)/j \leq b$ con $j = 1, ..., T$.

- $b_1 = (i-a) \cdot \frac{1}{2}(FVP_i + FVP_{i-1})$
- $b_2 = (b-j) \cdot \frac{1}{2}(FVP_{j+1} + FVP_j)$

Los valores del área parcial recorren el intervalo $[0,1]$ estando acotados superiormente por el área del rectángulo de base $b-a$ y altura 1. Es decir:

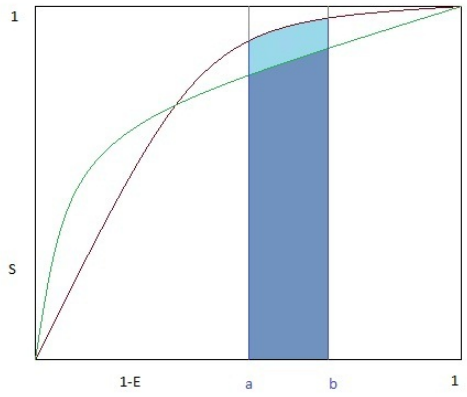
$$0 \leq \widehat{AUC}_{(a,b)} \leq b-a \leq 1$$

Por tanto, para dadas dos curvas hacer un estudio comparativo de sus áreas parciales habría que estudiar el valor de:

$$\widehat{AUC}^{(1)}_{(a,b)} - \widehat{AUC}^{(2)}_{(a,b)}$$

Siendo $\widehat{AUC}^{(1)}_{(a,b)}$ el área parcial entre las abcisas ($x=a$, $x=b$) de una de las dos curva y $\widehat{AUC}^{(2)}_{(a,b)}$ lo mismo la de la segunda curva. Sus valores recorren el intervalo $[0, \frac{1}{2}]$, correspondiendo el valor $\frac{1}{2}$ al caso límite en el que ($a=0$, $b=1$), una de las curvas sea la diagonal del cuadrado $[0,1]\times[0,1]$ y la otra se aproxime a los lados izquierdo y superior del mismo cuadrado.

Figura 3.8.2: Áreas parciales de dos curvas ROC.



Observación. No es un valor de interés la diferencia de área parciales en distintos intervalos.

3.9. Comparación de pruebas

Hemos visto que dadas dos pruebas es mejor, en el sentido discriminante, la que mayor área tenga y, si fuesen iguales, podría ser porque son la misma curva ROC o porque una nos ofrece una prueba más sensible y otra más específica. Sin

embargo, no podemos olvidar que nuestros datos siempre van a tener la variabilidad inevitable del muestreo. Por tanto, se nos presenta un nuevo problema, ¿una prueba tiene mayor área que otra única y exclusivamente por tener mayor capacidad discriminante o es debido a la variabilidad de la muestra? lo que nos lleva a plantearnos el siguiente contraste de hipótesis:

$$\begin{cases} H_0: & AUC_1 = AUC_2 \\ H_1: & AUC_1 \neq AUC_2 \end{cases}$$

Hanley y McNeil[11] proponen el siguiente estadístico para resolver este contraste:

$$z = \frac{(\widehat{AUC}_1 - \widehat{AUC}_2)}{SD(\widehat{AUC}_1 - \widehat{AUC}_2)}$$

Siendo SD la desviación estándar. Dicho estadístico sigue una distribución $N(0,1)$ bajo la hipótesis nula de modo que para un nivel de significación α se rechaza la aleatoriedad de nuestra prueba si $|z| > z_{\alpha/2}$, donde $z_{\alpha/2}$es el cuantil de dicha distribución que cumple $\Phi(z_{\alpha/2}) = 1 - \alpha/2$. En caso de $\alpha = 0{,}05$ rechazaríamos si $z > 1{,}96$.
Veámos el cálculo del denominador:

$$SD(\widehat{AUC}_1 - \widehat{AUC}_2) =$$

$$= \sqrt{\hat{V}(\widehat{AUC}_1) + \hat{V}(\widehat{AUC}_2) - 2 \cdot r \cdot SD(\widehat{AUC}_1) \cdot SD(\widehat{AUC}_2)}$$

Donde r es el coeficiente de correlación entre ambas áreas, siendo su expresión:

$$r = \frac{cov(\widehat{AUC}_1, \widehat{AUC}_2)}{SD(\widehat{AUC}_1) \cdot SD(\widehat{AUC}_2)}$$

Cuyo valor se obtiene a través de una tabla de correlaciones calculada por Hanley y McNeil [11] y unos valores promedios de las correlaciones en el grupo con presencia del evento, $D = 1$ y con ausencia, $D = 0$ y, por otro lado de la área calculadas también en ámbos grupos por separado.

Hay que tener en cuenta que este contraste es para rechazar o no la igualdad de áreas y, como hemos visto, dos curvas ROC muy diferentes podrían tener igual área. Por tanto, en caso de aceptar la iguadad, habría que acompañar el resultado con un examen visual de las curvas.

Contraste para la aleatoriedad de la prueba.

Una vez obtenido el valor del área bajo la curva para una prueba es recomendable someter ésta al contraste:

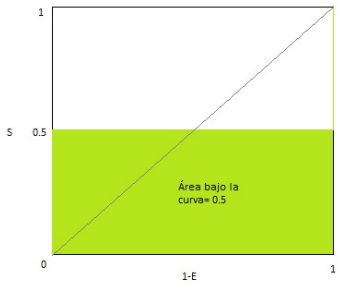
$$\begin{cases} H_0: & AUC = 0{,}5 \\ H_1: & AUC \neq 0{,}5 \end{cases}$$

Es decir, someter a contraste la posibilidad de que la variable de decisión sea aleatoria y no nos haya salido reflejado en el cálculo del AUC por la variabilidad de la muestra. El estadístico que resuelve este contraste viene dado por:

$$z = \frac{\widehat{AUC} - 0{,}5}{SD(\widehat{AUC})}$$

Que al igual que el estadístico anterior sigue una distribución $N(0,1)$ bajo hipótesis nula y, al igual que el contraste anterior, hay que acompañarlo de un examen visual de la curva pues se podría presentar el siguiente, aunque poco común, ejemplo:

Figura 3.9.1: Curva ROC con AUC de 0.5



La figura 3.9.1 representa el caso presentado por Hilden[45]quien dictó que, a pesar de su baja área, era una prueba válida. Los resultados de la prueba comprendidos en el intervalo $(80, 120)$ correspondían a individuos sanos, mientras que la mitad de los enfermos obtenían resultados por debajo de 80 y la otra mitad por encima de 120.

Sin embargo, aplicando la siguiente transformación monótona a la variable: $x' = |x - 100|$ se obtiene que sólo los valores por encima 20 corresponden a enfermos. Además, quedaría una curva de máxima área, es decir, $AUC = 1$.

Capítulo 4

Elección del punto de corte

Una vez que ya hemos dibujado la curva ROC y hemos visto, mediante su área, si tiene un apropiado nivel discriminante, toca escoger *punto de corte* o *valor umbral*. Una propiedad deseable para el mismo sería que tuviese asociada una alta sensibilidad y especificidad.

Esto puede conseguirse elegiendo el valor de corte que más alto índice de Youden tenga (véase Sección 3.2). El problema de este método es que estaríamos escogiendo la sensibilidad y especificidad más alta de manera conjunta. En consecuencia, necesitamos métodos alternativos que nos permitan ser más laxos con, por ejemplo, la especificidad si queremos una prueba muy sensible (prevención de enfermedades) o viceversa (tamizaje de enfermedades). Para ello, tenemos dos caminos: escoger el punto de mínima distancia al vértice $(0, 1)$ o minizar los costes de los resultados erróneos.

Lema 1. *Dada una curva ROC, el punto de corte correspondiente al par* $(1 - E, S)$ *más cercano al* $(0, 1)$ *es aquel cuya recta tangente tiene por pendiente*

$$m = \frac{p \cdot P_r(falsos\,positivos)}{(1-p) \cdot P_r(falsos\,negativos)}$$

siendo p *la prevalencia del evento en la población.*

Demostración. Sea $y = f(x)$ la función que define la curva ROC con $(x, y) = (1-E, S)$. La distancia de un punto cualquiera de la curva (x, y) al vértice $(0, 1)$ es una función de x que puede ser expresada como sigue:

$$dist(x) = x^2 + (y-1)^2 = x^2 + [f(x) - 1]^2$$

Buscamos el mínimo:

$$2x + 2[f(x) - 1]f'(x) = 0 \Longrightarrow f'(x) = \frac{x}{1 - f(x)}$$

Es decir, será el punto:

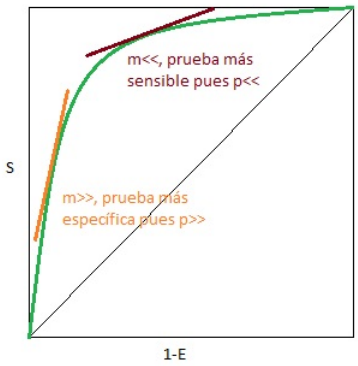
$$x/f'(x) = \frac{1-E}{1-S} = \frac{1-P_r(Y=0|D=0)}{1-P_r(Y=1|D=1)} = \frac{P_r(Y=1|D=0)}{P_r(Y=0|D=1)} =$$

$$\frac{P_r(Y=1\bigcap D=0)}{P_r(D=0)} \cdot \frac{P_r(D=1)}{P_r(Y=0\bigcap D=1)} = \frac{p \cdot P_r(falsos\, positivos)}{(1-p)\cdot P_r(falsos\, negativos)}$$

$$m = \frac{p \cdot P_r(falsos\, positivos)}{(1-p)\cdot P_r(falsos\, negativos)}$$

Por tanto el mínimo se alcanzará en el punto cuya pendiente sea m. □

Figura 4.0.1: Significado de pendientes en ejemplo de curva ROC bajo Lema 1



Observación. Si la prevalencia fuese alta, la fracción $\frac{p}{1-p}$ también lo sería y, a su vez, m. El *lema* 1 nos llevaría a puntos de la parte izquierda de la curva donde la pendiente es grande y tendríamos por tanto una prueba más específica y menos sensible que disminuiría los F_+.

Por el contrario, si la prevalencia fuese baja nos llevaría a puntos de la parte derecha de la curva, disminuyendo los F_- y dándonos una prueba más sensible y menos específica.

Lema 2. *Sea una curva ROC asociada a un diagnóstico sobre un evento y considérense los costes, previamente fijados, de los resultados erróneos, entonces el punto de corte correspondiente al par* $(1-E, S)$ *que minimiza dichos costes es aquel cuya recta tangente tiene por pendiente:*

$$m = \frac{costes\, falsos\, positivos \cdot (1-p)}{costes\, falsos\, negativos \cdot p}$$

siendo p *la prevalencia del evento.*

Demostración. Se define el coste medio esperado del uso de un diagnóstico como:

$$C_{esp} = C_0 + C_{V_+} \cdot P_r(V_+) + C_{V_-} \cdot P_r(V_-) + C_{F_+} \cdot P_r(F_+) + C_{F_-} \cdot P_r(F_-)$$

siendo:

$$\begin{cases} C_0 & Coste\,base \\ C_{V_+} & Coste\,de\,verdaderos\,positivos \\ C_{V_-} & Coste\,de\,verdaderos\,negativos \\ C_{F_+} & Coste\,de\,falsos\,positivos \\ C_{F_-} & Coste\,de\,falsos\,negativos \end{cases}$$

y siendo:

$$\begin{cases} P_r(V_+) = P_r(Enfermo) \cdot P_r(Y = 1|Enfermo) = p \cdot S & prevalencia \times sensibilidad \\ P_r(V_-) = P_r(Sano) \cdot P_r(Y = 0|Sano) = (1-p) \cdot E & (1 - prevalencia) \times especificidad \\ P_r(F_+) = P_r(Sano) \cdot P_r(Y = 1|Sano) = (1-p) \cdot (1-E) & (1 - prevalencia) \times (1 - especificidad) \\ P_r(F_-) = P_r(Enfermo) \cdot P_r(Y = 0|Enfermo) = p \cdot (1-S) & prevalencia \times (1 - sensibilidad) \end{cases}$$

sustituyendo:

$$C_{esp} = C_0 + C_{V_+} \cdot p \cdot S + C_{V_-} \cdot (1-p) \cdot E + C_{F_+} \cdot (1-p) \cdot (1-E) + C_{F_-} \cdot p \cdot (1-S) =$$

$$= C_0 + C_{V_+} \cdot p \cdot f(x) + C_{V_-} \cdot (1-p) \cdot (1-x) + C_{F_+} \cdot (1-p) \cdot x + C_{F_-} \cdot p \cdot (1-f(x)) =$$

reagrupando:

$$= C_0 + C_{V_-} \cdot (1-p) + C_{F_-} \cdot p + p \cdot (C_{V_+} - C_{F_-}) \cdot f(x) + (1-p) \cdot (C_{F_+} - C_{V_-}) \cdot x$$

Buscamos el mínimo de ésta función, derivando e igualando a 0:

$$p \cdot (C_{V_+} - C_{F_-}) \cdot f'(x) + (1-p) \cdot (C_{F_+} - C_{V_-}) = 0$$

$$f'(x) = \frac{(C_{F_+} - C_{V_-}) \cdot (1-p)}{-(C_{V_+} - C_{F_-}) \cdot p}$$

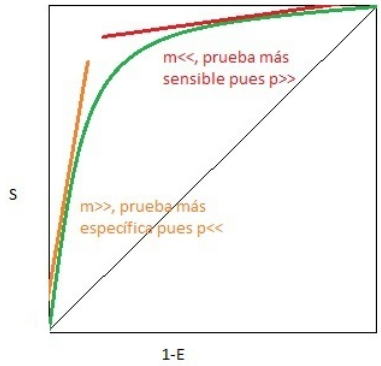
Ahora bien, si consideramos que los costes de los resultados acertados son nulos, es decir, $C_{V_-} = 0 = C_{V_+}$ nos queda:

$$f'(x) = \frac{C_{F_+} \cdot (1-p)}{C_{F_-} \cdot p}$$

$$m = \frac{C_{F_+} \cdot (1-p)}{C_{F_-} \cdot p}$$

El mínimo se alcanzará en el punto que tenga por pendiente m. □

Figura 4.0.2: Significado de pendientes en ejemplo de curva ROC bajo Lema 2



Observación. Cuando la prevalencia es baja, la fracción $\frac{(1-p)}{p}$ será grande. Por tanto, el *lema* 2 nos estará llevando a puntos con baja sensibilidad y alta especificidad, esto disminuye los F_+. Además, ocurre igual si $C_{F_+} \ggg C_{F_-}$ pues la pendiente, m, será grande y nos llevará a la parte izquierda de la curva.

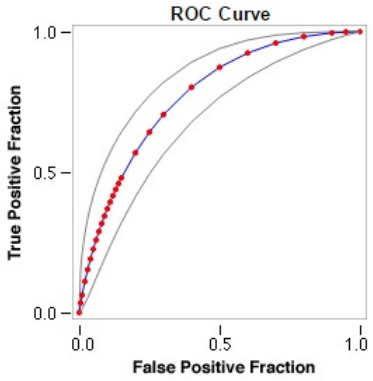
Por el contrario, cuando la prevalencia sea alta, la fracción $\frac{(1-p)}{p}$ será pequeña y obtendremos una pendiente leve. Lo que se traduce en escoger un punto con alta sensibilidad y baja especificidad, haciendo así que los F_- disminuyan. Ésto también ocurre si $C_{F_+} \lll C_{F_-}$ pues hará disminuir a la derivada y por tanto a m, llevándonos a puntos de la parte derecha de la curva.

Capítulo 5

Softwares estadísticos para las curvas ROC

Existen varios programas con los que podemos aplicar el análisis ROC a una base de datos: *Rockit*, *Metz ROC, AccuROC* y *ROC Analysis*. Este último es online y nos ofrece bajo unos datos generados la gráfica de la curva ROC, los pares de puntos (FFP, FVP) que la dibujan, el área bajo la curva (tanto la estimada por el modelo binormal como la calculada por el estadistico de Wilcoxon), los parámetros a y b estimados del modelo binormal y el intervalo de confianza para FFP y FVP para cada punto de corte.

Figura 5.0.1: Curva ROC en ROC Analysis.[25]



En este capítulo nos centraremos en dos de los paquetes estadísticos más usados *IBM SPSS Statistic* y *R: The R Project for Statistical Computing*. Para ello, vamos a plantearnos encontrar un *biomarcador* que nos diagnostique si un individuo padece o no la enfermedad *cáncer de páncrea*s. Contaremos con los datos[27] de las siguientes variables:

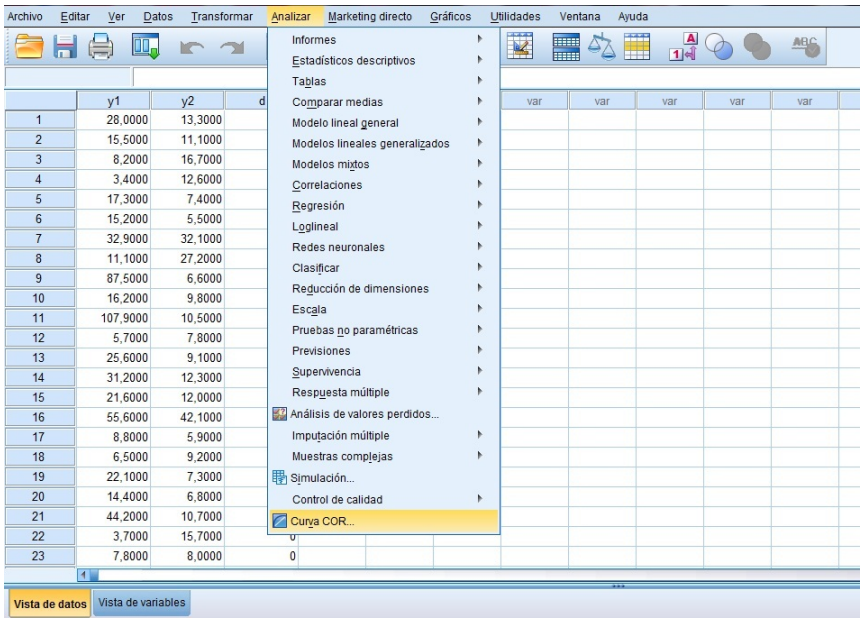
- Variable d: es el estado real de cada paciente. Toma el valor 0 cuando el paciente está sano y el valor 1 cuando padece la enfermedad.
- Variable y_1: corresponde al biomarcador $CA\,19-9$, es una variable de tipo continua y es la primera prueba que queremos validar.
- Variable y_2: es el biomarcador $CA\,125$, al igual que y_1 es una variable continua y será la segunda prueba a validar.

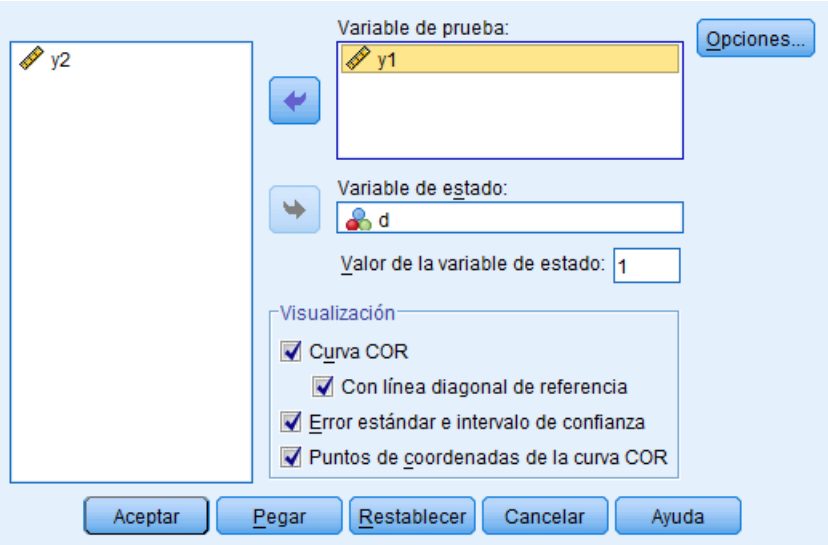
5.1. Análisis de curvas ROC en SPSS

Al abrir el documento en *Vista de Variables* tenemos:

	Nombre	Tipo	Anch...	Deci...	...	...	...	Co...	Alinea...	Medida	Rol
1	y1	Num...	8	4				8	Der...	Escala	Entrada
2	y2	Num...	7	4				8	Der...	Escala	Entrada
3	d	Num...	1	0				8	Der...	Nominal	Entrada

La opción de analizar una curva ROC en encuentra en el menú desplegable *Analizar/Curva COR*:

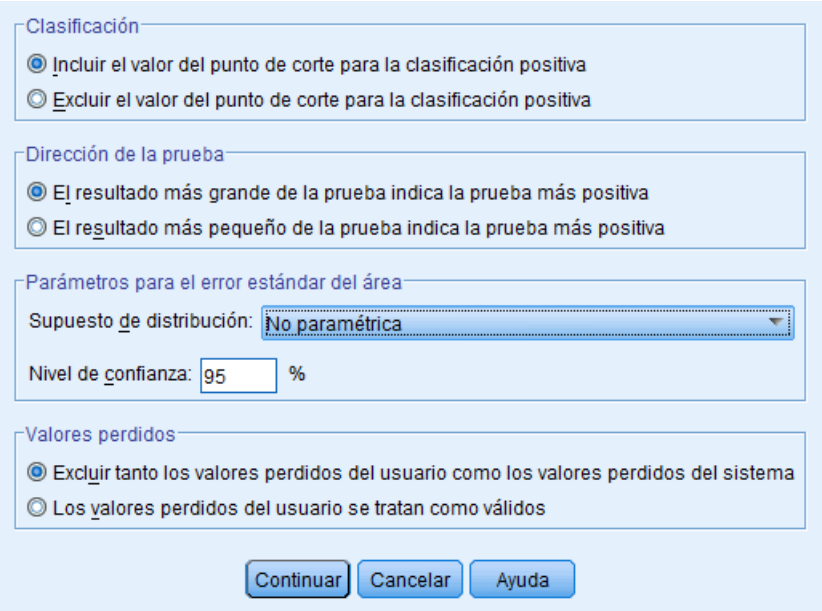




En *Variable de prueba* añadimos la o las variables correspondientes a las pruebas que estemos evaluando, éstas deben ser *numéricas*. En *Variable de estado* debe ir aquella que etiquete a cada individuo con $\{0, 1\}$ en función de su estado, además hay que especificar en *Valor de la variable estado* cuál de esos dos valores representa el evento de estudio.

En la sección de Visualización incluimos que nos grafique la curva con la diagonal del cuadrado $[0, 1] \times [0, 1]$, es decir la curva ROC correspondiente a la prueba aleatoria. Además, incluimos que nos devuelva la estimación de la varianza y el intervalo de confianza para el área bajo la curva. Por último, obtendremos una tabla con todos los pares de sensibilidad y especificidad usados para construir la curva.

Entrando en *Opciones* aparece una nueva ventana. En su apartado de *Clasificación* podemos decidir si al hacer la dicotomización de los valores de la prueba incluimos el punto de corte en los calificados por la prueba como enfermos o como sanos. En *Dirección de la prueba* le hacemos saber si los resultados más altos de la prueba corresponden a enfermos o sanos. Esta opción sirve para invertir la positividad de la curva y convertirla en cóncava cuando salga convexa. Por defecto, la curva se construye mediante un método no paramétrico pero podemos hacerla también suponiendo que los datos siguen una distribución *exponencial binegativa*, más usada en estudios de Análisis de Supervivencia. También podemos variar el *nivel de confianza* para el cálculo del intervalo de confianza del área. Finalmente podemos decidir si excluir o no aquellos pacientes a los que les falte algún dato.



Con las opciones marcadas en las figuras anteriores se obtiene la siguiente salida:

Resumen de procesamiento de casos

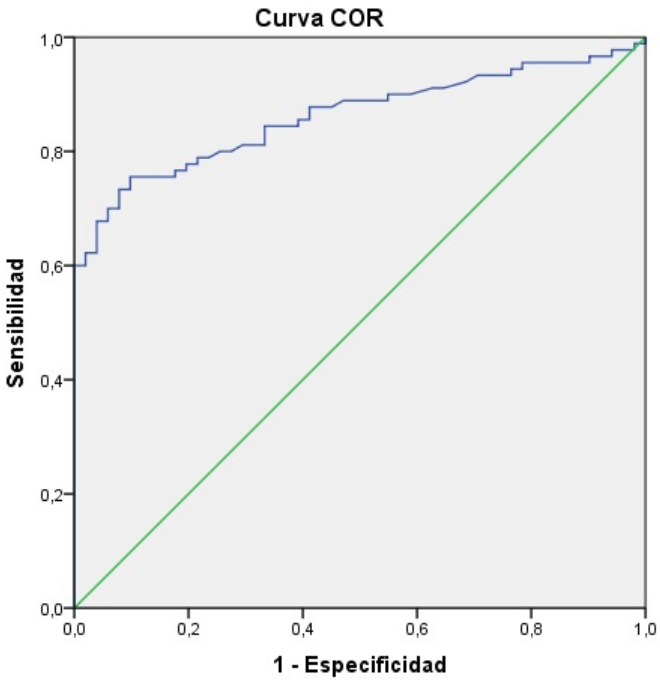
d	N válido (por lista)
Positivo[a]	90
Negativo	51

Los valores más grandes de la(s) variable(s) de resultado de prueba indican una prueba mayor para un estado real positivo.

a. El estado real positivo es 1.

Nuestra muestra se compone de un total de 141 pacientes de los cuales 90 padecen cáncer de páncreas y 51 son sanos o casos de control. La estimación de la curva ROC para dicho evento bajo ésta muestra es:

Figura 5.1.1: Curva ROC para la variable y_1



Área bajo la curva

Variable(s) de resultado de prueba: y1

Área	Error estándar[a]	Significación asintótica[b]	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
,861	,031	,000	,802	,921

La(s) variable(s) de resultado de prueba: y1 tiene, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Las estadísticas podrían estar sesgadas.

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

El valor estimado para el área es del 86,1 %, su intervalo de confianza no disminuye del 80 % y sobrepasa el 90 % luego podemos determinar que este biomarcador tiene una buena capacidad discriminante. En *Significación asintótica* encontramos un *p-valor* de 0 para el contraste $H_0 : AUC = 0{,}5$, es decir, H_0 : *La prueba es aleatoria*, hipótesis que rechazamos. En la casilla de error estándar obtenemos un bajo valor para la estimación de la varianza del área, esto es un indicativo de que se aproxima al verdadero valor del área que tendría la real curva ROC. Por último, nos devuelve la tabla con cada coordenada de la curva indicando a qué punto de corte corresponde. La siguiente imagen es una

extracción de dicha tabla, para valores umbrales entorno a 13 obtendríamos una prueba con una sensibilidad de más del 80 % y una especificidad de poco menos del mismo valor. Para la elección exacta del punto de corte habría que sopesar si queremos una prueba más sensible, es decir, preventiva, ó una más específica, es decir, tamizaje de emfermedades.

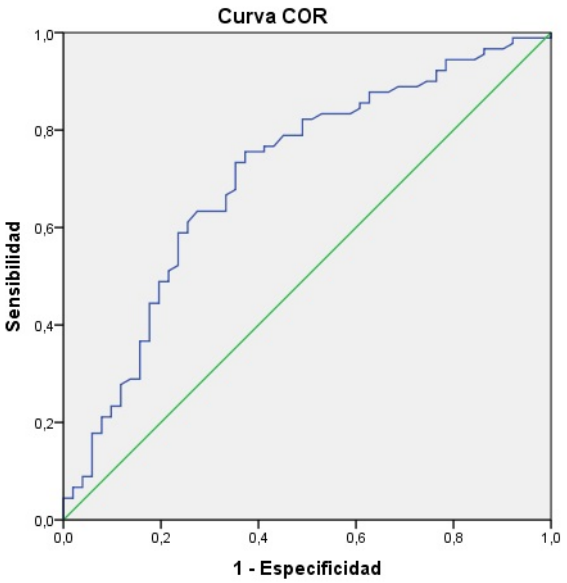
Coordenadas de la curva

Variable(s) de resultado de prueba: y1

Positivo si es mayor o igual que[a]	Sensibilidad	1 - Especificidad
1,400000	1,000	1,000
2,900000	,989	1,000
3,500000	,989	,980
3,625000	,978	,980
3,675000	,978	,961
3,800000	,978	,941
3,950000	,967	,941
8,625000	,889	,549
8,850000	,889	,529
9,450000	,889	,510
10,100000	,889	,490
10,300000	,889	,471
10,750000	,878	,451
11,300000	,878	,431
12,150000	,878	,412
13,600000	,856	,412
14,550000	,856	,392
14,950000	,844	,392
15,300000	,844	,373
15,450000	,844	,353
15,550000	,844	,333
15,650000	,833	,333
15,950000	,811	,333
16,750000	,811	,314

A continuación se muestra el mismo análisis para la variable y_2:

Figura 5.1.2: Curva ROC para la variable y_2



Área bajo la curva

Variable(s) de resultado de prueba: y2

Área	Error estándar[a]	Significación asintótica[b]	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
,706	,047	,000	,614	,797

La(s) variable(s) de resultado de prueba: y2 tiene, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Las estadísticas podrían estar sesgadas.

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

Tiene un área del 70,6 %, menor que el 86,1 % que alcanzaba y_1, su intervalo de confianza baja casi hasta el 60 % y no alcanza el 80 %. La estimación de su varianza es algo mayor que la de y_1, sin embargo, no es una prueba aleatoria ya que obtenemos un p-valor muy significativo. Por tánto, a la vista de éstos resultados podemos apuntar a que aunque la variable y_2 pueda ser útil para algunos propósitos, parece que es y_1 quien mejor detecta qué individuo padece o no cáncer de páncreas.

Por si, bajo alguna situación clínica, fuése más conveniente realizar la prueba y_2 veámos los mejores puntos de corte a escoger:

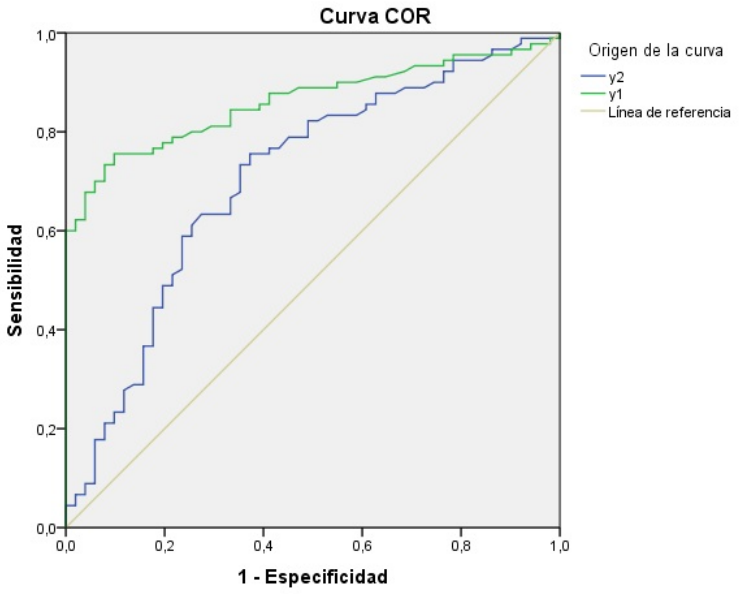
Coordenadas de la curva

Variable(s) de resultado de prueba: y2

Positivo si es mayor o igual que[a]	Sensibilidad	1 - Especificidad
2,700000	1,000	1,000
4,600000	,989	1,000
5,550000	,989	,980
5,750000	,989	,961
6,200000	,989	,922
6,550000	,978	,922
6,700000	,967	,902
10,450000	,844	,608
10,550000	,833	,588
10,650000	,833	,569
10,900000	,833	,549
11,150000	,833	,529
11,300000	,822	,510
11,500000	,822	,490
11,650000	,811	,490
11,850000	,789	,490
12,050000	,789	,451
12,200000	,767	,431
12,400000	,767	,412
12,550000	,756	,412
12,800000	,756	,373
13,100000	,744	,373
13,250000	,733	,373
13,700000	,733	,353

Escogiendo un punto de corte entrono a 11 mantendríamos la sensibilidad al 80 % pero nos bajaría la especificidad casi al 50 %. Si mantenemos el umbral alrededor de 13, el escogido con y_1, obtendremos una especificidad y espeficidad del 70 %. Una utilidad de ésta prueba sería con la finalidad de obtener una prueba muy específica. Veamos que para valores umbrales muy altos ambas curvas toman coordenadas muy cercanas.

Figura 5.1.3: Curvas ROC para las variables y_1 e y_2.



Coordenadas de la curva

Variable(s) de resultado de prueba: y1

Positivo si es mayor o igual que[a]	Sensibilidad	1 - Especificidad
24,650000	,778	,216
26,450000	,778	,196
27,650000	,767	,196
28,300000	,767	,176
29,900000	,756	,176

Coordenadas de la curva

Variable(s) de resultado de prueba: y2

Positivo si es mayor o igual que[a]	Sensibilidad	1 - Especificidad
22,350000	,489	,216
23,650000	,489	,196
25,250000	,478	,196
26,133350	,467	,196
26,633350	,456	,196

Escogiendo un valor umbral alrededor de 26 obtenemos para la prueba y_1 una sensibilidad y especificidad del casi 80 % mientras que en y_2 mantendríamos la especificidad al 80 % pero la sesibilidad quedaría prácticamente al 50 %. Por lo que definitivamente es y_1 la mejor variable de decisión de las dos para discernir entre pacientes con o sin cáncer de práncreas.

5.2. Análisis de la curva ROC en R

El paquete más utilizado para la aplicación de las curvas ROC es el denominado *pROC* Display and Analyze ROC curves[28]. Este permite representar curvas ROC, calcular el área bajo ellas, comparar pruebas, calcular intervalos de confianza y calcular áreas parciales. De modo que las primeras líneas de código en nuestro script serán:

```
install.packages("pROC")
library(pROC)
```

A continuación, se ha de leer los datos que tenemos previamente guardados en el archivo *Panc.txt*, para ello usamos la función *read.table*():

```
datos<- read.table(file="Panc.txt",header=TRUE)
```

Con las funciones *summary*() y *str*() obtenemos el siguiente resumen de los datos:

```
summary(datos)
str(datos)
names(datos)
```

```
> summary(datos)
       y1                 y2                d
 Min.   :    2.4   Min.   :   3.70   Min.   :0.0000
 1st Qu.:   10.0   1st Qu.:  10.50   1st Qu.:0.0000
 Median :   44.2   Median :  17.20   Median :1.0000
 Mean   : 1101.5   Mean   :  43.03   Mean   :0.6383
 3rd Qu.:  508.0   3rd Qu.:  35.00   3rd Qu.:1.0000
 Max.   :24000.0   Max.   :1024.00   Max.   :1.0000
> str(datos)
'data.frame':   141 obs. of  3 variables:
 $ y1: num  28 15.5 8.2 3.4 17.3 15.2 32.9 11.1 87.5 16.2 ...
 $ y2: num  13.3 11.1 16.7 12.6 7.4 5.5 32.1 27.2 6.6 9.8 ...
 $ d : int  0 0 0 0 0 0 0 0 0 0 ...
> names(datos)
[1] "y1" "y2" "d"
```

Vemos que y_1 toma un rango de valores más amplio que y_2. Para representar las curvas ROC necesitaremos poder llamar a cada variable por separado, por lo que definimos:

```
d<-datos[,3] #variable estado
y1<-datos[,1]#biomarcador CA 19-9
y2<-datos[,2]#biomarcador CA 125
```

La función que representa la curva ROC es *roc*(). Cuyos argumentos más relevantes son:

- La variable estado real d.
- La variable de decisión a estudiar y_1 ó y_2.

- Controls, cases: para introducir las variables anteriores como dos vectores ordendos donde para cada posición encontramos el estado y el valor del biomarcador de un mismo paciente.
- Percent=true/false: expresa los resultados en porcentaje o en fracción.
- Na.rm=true/false: elimina o no aquellos pacientes a los que le falte algún dato.
- Direction: cambia el grupo de casos por el de controles, es decir, este argumento nos sirve para cuando los enfermos tienen menor calificacion en la prueba que los sanos y necesitemos invertir la positividad de la curva para que salga cóncava y no convexa.
- Algorithm=1/2/3/4: Algoritmo para calcular los estimadores de sensibilidad y especificidad. El que se usa por defecto es el 1, el 4 realiza los tres algoritmos anteriores y comprueba si dan o no los mismos resultados.
- Smooth=true/false: Suaviza o no la curvatura de la curva.
- Auc=true/false: Calcula o no el área bajo la curva.
- Ci=true/false: Obtiene el intervalo de confianza para el área bajo la curva.
- Plot=true/false: Devuelve o no el gráfico de la curva.
- Smooth.method: En caso de $Smooth = TRUE$ especifíca a qué densidades queremos ajustar tanto el grupo de enfermo como el de sanos. Por defecto ajusta a la binormal.

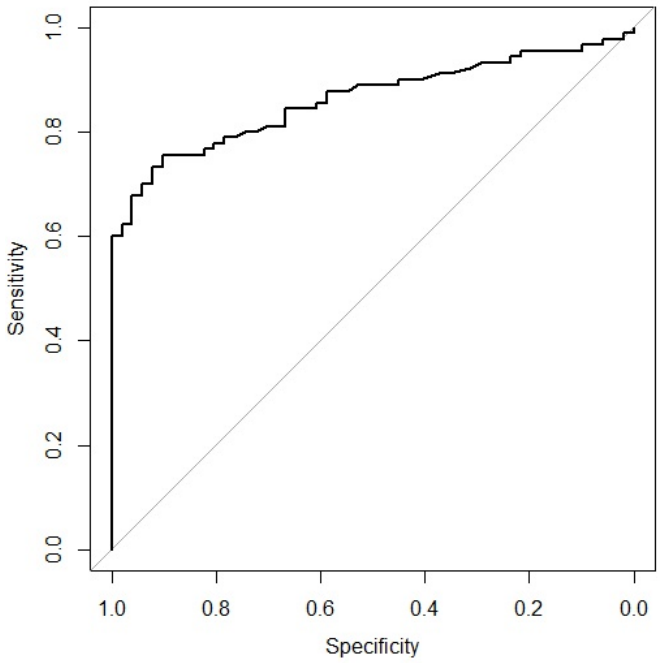
Por tanto, el código para calcular la curva ROC no paramétrica de la variable $y1$ en nuestro script quedaría:

```
ROCy1<-roc(d, y1, percent=FALSE, na.rm=TRUE,
direction=c("auto", "<", ">"),smooth=FALSE, auc=TRUE,
ci=TRUE, plot=TRUE)
ROCy1
```

Que nos devuelve:

```
Data: 51 controls < 90 cases.
Area under the curve: 0.8614
95% CI: 0.8015-0.9214 (DeLong)
```

Figura 5.2.1: Curva ROC no paramétrica de la variable y_1



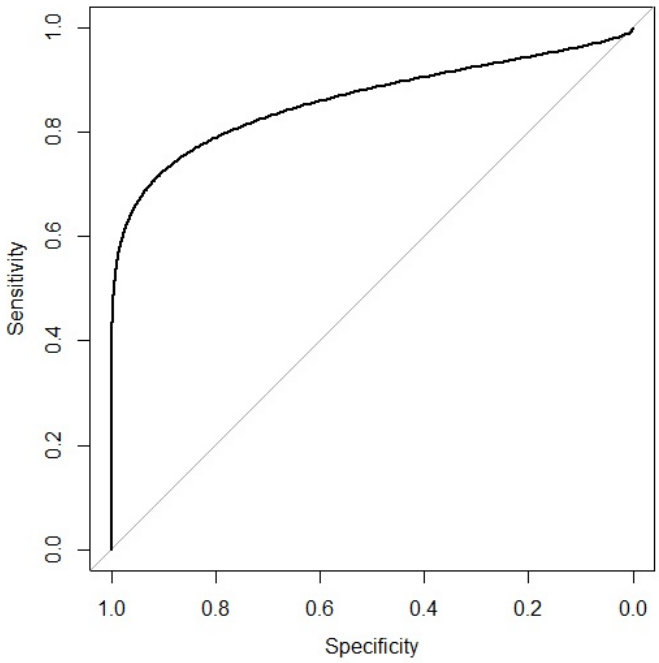
Cambiando el argumento *Smooth* obtenemos la curva ROC paramétrica de la variable y_1 donde las distribuciones de sanos y enfermos han sido ajustadas a una distribución binormal. Ttambién podemos ajustarla a una suavización Kernel usando $smooth.method = c(\text{"}density\text{"})$. Además, eliminando el argumento $ci.method = NULL$ calcula el intervalo de confianza para el área realizando 2000 muestras bootstrap.

```
ROCy1Smooth<-roc(d, y1, percent=FALSE, na.rm=TRUE,
direction=c("auto", "<", ">"), smooth=TRUE, auc=TRUE, ci=TRUE,
plot=TRUE, smooth.method="binormal", density=NULL)
ROCy1Smooth
```

Nos devuelve:

```
Data: y1 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: binormal
Area under the curve: 0.8613
95% CI: 0.7972-0.917 (2000 stratified bootstrap replicates)
```

Figura 5.2.2: Curva ROC paramétrica de la variable y_1



Por otra parte, también podemos obtener directamente el área mediante la función *auc*(), su argumento es el nombre de la curva aunque también se le pueden incluir argumentos como $partial.auc = TRUE$ o $partial.auc.focus = c(\text{"}specificity\text{"},\text{"}sensitivity\text{"})$ para calcular áreas parciales. Especialmente útil para curvas diferentes con áreas muy próximas. Al ejecutarlo en nuestro script observamos que las estimaciones del área por los método no paramétrico y paramétrico son prácticamente iguales.

```
> auc(ROCy1)
Area under the curve: 0.8614
> auc(ROCy1Smooth)
Area under the curve: 0.8613
```

Con la función *Smooth*() podemos obtener la suavización de nuestra curva mediante los métodos: binormal, density (el correspondiente al método Kernel), fitdistr (incluida en el paquete MASS), logcondens y logcondens.smooth (incluidos éstos dos últimos en el paquete logcondens). Pueden darse considerables cambios en las estimaciónes del área según el método utilizado, esto dependerá de la precisión de las estimaciones y tamaños muestrales.

```
> smooth(ROCy1, method=c("binormal"))

Call:
smooth.roc(roc = ROCy1, method = c("binormal"))

Data: y1 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: binormal
Area under the curve: 0.8613

> smooth(ROCy1, method=c("density"), n=512, bw = "nrd0")

Call:
smooth.roc(roc = ROCy1, method = c("density"), n = 512, bw = "nrd0")

Data: y1 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: density (bandwidth: nrd0; adjust: 1)
Area under the curve: 0.7913

> smooth(ROCy1, method=c("fitdistr"), n=512, bw = "nrd0")

Call:
smooth.roc(roc = ROCy1, method = c("fitdistr"), n = 512, bw = "nrd0")

Data: y1 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: fitdistr
Area under the curve: 0.9946


> smooth(ROCy1, method=c("logcondens"), n=512, bw = "nrd0")

Call:
smooth.roc(roc = ROCy1, method = c("logcondens"), n = 512, bw = "nrd0")

Data: y1 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: logcondens
Area under the curve: 0.9904

> smooth(ROCy1, method=c("logcondens.smooth"), n=512, bw = "nrd0")

Call:
smooth.roc(roc = ROCy1, method = c("logcondens.smooth"), n = 512,

Data: y1 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: logcondens.smooth
Area under the curve: 0.6725
```

Para obtener directamente el intervalo de confianza del área de la curva podemos usar las funciones *ci*() y *ci.auc*(). A ambas se les puede añadir como argumento el nombre del elemento ROC que hayamos creado o los datos en dos vectores, uno correspondiente a sanos y otro a enfermos. Además, con el argumento *method* se le puede especificar si se quiere calcular con el estadístico DeLong o mediante el método bootstrap.

```
> ci(ROCy1)
95% CI: 0.8015-0.9214 (DeLong)
> ci.auc(ROCy1)
95% CI: 0.8015-0.9214 (DeLong)
```

Otras funciones de interés son *ci.sp*() y *ci.se*(), que permiten calcular el intervalo de confianza de la especificidad y sensibilidad para cada valor estimado de sensibilidad y especificidad respectivamente. Sus argumentos son similares a las funciones anteriores, si no se le especifica lo contrario realizará 2000 muestras bootstrap. En la salida obtenemos una tabla de cuatro columnas, la primera corresponde al recorrido de valores de sensibilidad o especificidad llendo desde 0 hasta 1 sumando 0,1, la segunda y la cuarta corresponden a los extremos de los intervalos y la tercera una media de los mismos.

```
> ci.se(ROCy1)
95% CI (2000 stratified bootstrap replicates):
  sp se.low se.median se.high
 0.0 1.0000    1.0000  1.0000
 0.1 0.9111    0.9667  1.0000
 0.2 0.8889    0.9444  0.9889
 0.3 0.8556    0.9287  0.9778
 0.4 0.8333    0.9089  0.9667
 0.5 0.8083    0.8889  0.9556
 0.6 0.7667    0.8556  0.9333
 0.7 0.7222    0.8222  0.9111
 0.8 0.6889    0.7778  0.8667
 0.9 0.6222    0.7444  0.8444
 1.0 0.5111    0.6111  0.7444

> ci.sp(ROCy1)
95% CI (2000 stratified bootstrap replicates):
  se sp.low sp.median sp.high
 0.0 1.0000    1.0000  1.0000
 0.1 1.0000    1.0000  1.0000
 0.2 1.0000    1.0000  1.0000
 0.3 1.0000    1.0000  1.0000
 0.4 1.0000    1.0000  1.0000
 0.5 1.0000    1.0000  1.0000
 0.6 0.9020    1.0000  1.0000
 0.7 0.7451    0.9412  1.0000
 0.8 0.5294    0.7647  0.9608
 0.9 0.1765    0.4706  0.7255
 1.0 0.0000    0.0000  0.1373
```

La función *coords*() devuelve los valores de sensibilidad y especificidad para un valor umbral, sus argumentos son el elemento ROC y dicho valor umbral. Si mantenemos el punto de corte en 26 obtendríamos una sensibilidad de más del 70 % y una e-specificidad de más del 80 %. Si quisieramos una sensibilidad mayor sería a costa de bajar la especificidad a casi el 50 % y el punto de corte bajaría a la mitad, 13.

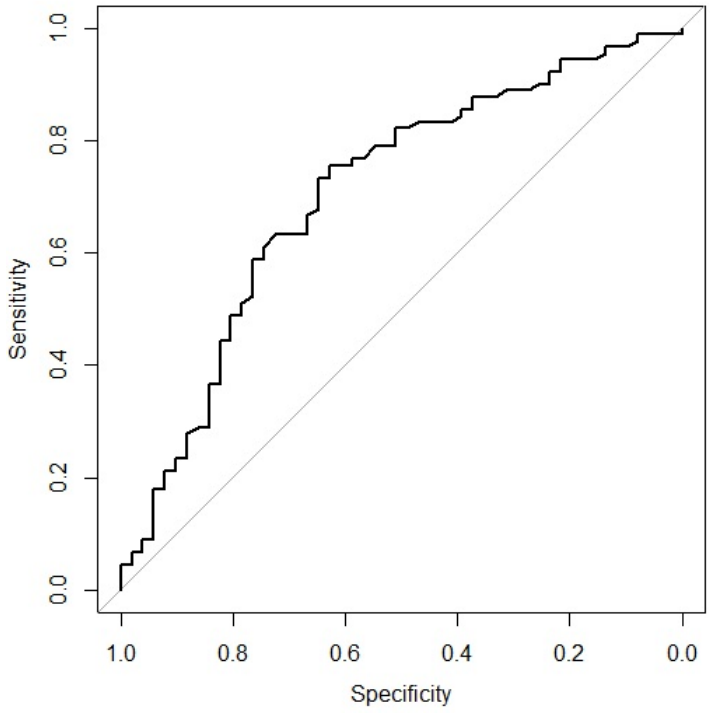
```
> coords(ROCy1,13)
  threshold specificity sensitivity
 13.0000000   0.5882353   0.8555556
> coords(ROCy1,26)
  threshold specificity sensitivity
 26.0000000   0.8039216   0.7777778
```

Apliquemos ahora el análisis anterior a la variable y_2 y comparemos resultados. Su curva ROC no paramétrica tiene un área de 0,7056, menor de 0,8614, el área de la variable y_1.

```
ROCy2<-roc(d, y2, percent=FALSE, na.rm=TRUE,
direction=c("auto", "<", ">"),smooth=FALSE,
auc=TRUE, ci=TRUE, plot=TRUE)
ROCy2
```

```
Data: 51 controls < 90 cases.
Area under the curve: 0.7056
95% CI: 0.6138-0.7973 (DeLong)
```

Figura 5.2.3: Curva ROC no paramétrica de la variable $y2$

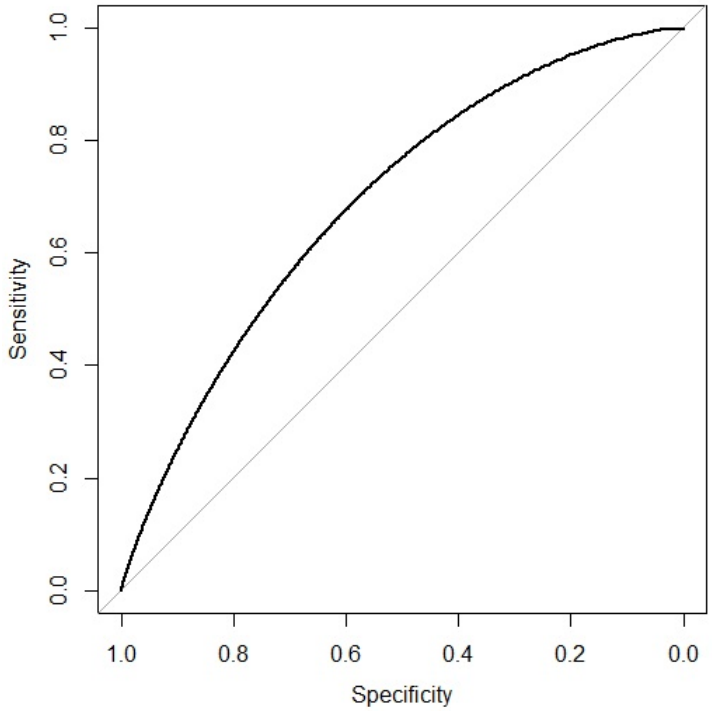


De igual manera que con y_1, mediante $smooth = TRUE$ calculamos la curva paramétrica ajustada por la distribución binormal. Nos da un área de 0,6904, con un intervalo de confianza que baja al 60 % y apenas llega al 80 %.

```
ROCy2Smooth<-roc(d, y2, percent=FALSE, na.rm=TRUE,
direction=c("auto", "<", ">"), smooth=TRUE, auc=TRUE,
ci=TRUE,plot=TRUE, smooth.method="binormal", density=NULL)
ROCy2Smooth
```

```
Data: y2 in 51 controls (d 0) < 90 cases (d 1).
Smoothing: binormal
Area under the curve: 0.6904
95% CI: 0.6021-0.7846 (2000 stratified bootstrap replicates)
```

Figura 5.2.4: Curva ROC paramétrica de la variable y_2



En los intervalos de confianza para la especificidad y sensibilidad respectivamente podemos ver que para valores altos de sensibilidad obtendríamos valores bajos de especificidad y viceversa.

```
> ci.sp(ROCy2)
95% CI (2000 stratified bootstrap replicates):
  se sp.low sp.median sp.high
 0.0 1.0000    1.0000  1.0000
 0.1 0.8824    0.9608  1.0000
 0.2 0.8235    0.9216  0.9804
 0.3 0.7451    0.8627  0.9608
 0.4 0.7059    0.8235  0.9216
 0.5 0.6667    0.7843  0.9020
 0.6 0.5686    0.7255  0.8627
 0.7 0.4706    0.6471  0.8039
 0.8 0.3134    0.5294  0.7255
 0.9 0.1176    0.2941  0.5294
 1.0 0.0000    0.0000  0.1765

> ci.se(ROCy2)
95% CI (2000 stratified bootstrap replicates):
  sp  se.low se.median se.high
 0.0 1.00000   1.00000  1.0000
 0.1 0.91110   0.96670  1.0000
 0.2 0.84440   0.93330  0.9889
 0.3 0.80000   0.88890  0.9667
 0.4 0.75550   0.85560  0.9333
 0.5 0.68890   0.81110  0.9000
 0.6 0.57780   0.75560  0.8556
 0.7 0.43330   0.64440  0.8111
 0.8 0.22220   0.48890  0.7022
 0.9 0.06667   0.24440  0.5234
 1.0 0.01111   0.05556  0.2000
```

Con la función *coords*() vemos que si queremos una sensibilidad de más del 80 % será a costa de tener una especificidad del 45 %, ésto se da con un valor umbral de 11. Si por el contrario queremos una especificidad del 80 % será bajo una sensibilidad del 46 % y un valor umbral de 26.

```
> coords(ROCy2,11)
  threshold specificity sensitivity
 11.0000000   0.4509804   0.8333333
> coords(ROCy2,26)
  threshold specificity sensitivity
 26.0000000   0.8039216   0.4666667
```

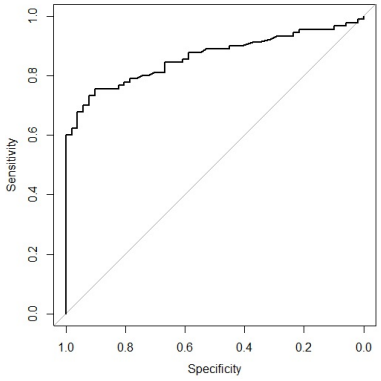
Por último, con la función *roc.test*() veamos si ambas variables conducen a unas curvas ROC significativamente distintas. Dicha función realiza un contraste de hipótesis donde $H_0 : AUC_{y_1} = AUC_{y_2}$. El p-valor es muy próximo a 0 luego rechazamos la hipótesis nula y aceptamos que son diferentes.

```
> roc.test(ROCy1,ROCy2)

        DeLong's test for two correlated ROC curves

data:  ROCy1 and ROCy2
Z = 2.7221, p-value = 0.006488
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
  0.8614379   0.7055556
```

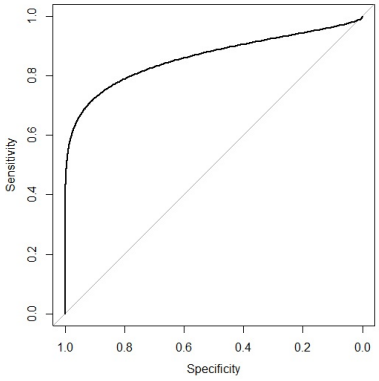
En conclusión, podemos determinar que el biomarcador $y1$ tiene mayor capacidad discriminante entre individuos con o sin cáncer de páncreas que el y_2. Para ello nos hemos basado en el valor del área, en un examen visual de ambas curvas y en los pares de sensibilidad y especificidad representados.

Figura 5.2.5: Comparativa de curvas ROC

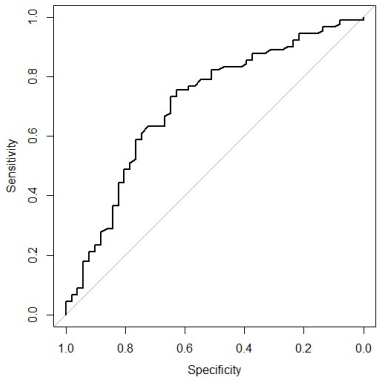
(a) ROC no paramétrica y_1



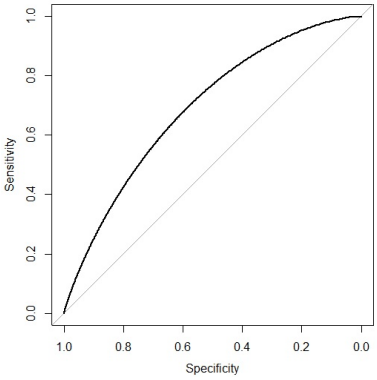
(b) ROC paramétrica y_1



(c) ROC no paramétrica y_2



(d) ROC paramétrica y_2



Capítulo 6

Aplicaciones

Desde su primera aplicación a la teoría de detección de señales, las curvas ROC han sido útiles en muy diversos campos como la psicología, la economía, el aprendizaje automático, la meteorología y la medicina, siendo ésta última especialmente predominante. Swets y Pickett[46] alegan que las razones que hacen atractivo el uso del análisis ROC son: los índices de exactitud intrínseca que lleva asociados, las estimaciones muestrales existentes de sensibilidad y especificidad y, por último, la libertad de variar el punto de corte para casos en que los resultados falsos negativos conlleven más costes que los positivos o viceversa. A su vez, cumple con los propósitos que según Sox et al[47], una herramienta de clasificación debería cumplir. A saber: indicar de manera fiable el estado en el que se encuentre un individuo y ser útil la información que proporcione para determinar, en el caso sanitario, un diagnóstico apropiado[16].
En lo que sigue de capítulo recogeremos algunas de las numerosas aplicaciones que se han hecho de las curvas ROC en cada uno de los campos mencionados.

Medicina

- Erkel y Pattynama (1998): '*Receiver-Operating Characteristic (ROC) analysis: basic principles and applications in radiology*'[70]: En este artículo se trata la construcción de curvas ROC y los diferentes tipos de variables que, en un entorno médico, se pueden encontrar, a saber: continuas (por ejemplo la medida de una lesión), escala categórica (los grados de gravedad de la estenosis de la arteria renal) o datos cualitativos (localización y avance de una lesión). Además, alegan que la elección del punto de corte para curvas suavizadas aun cuando los datos eran cualitativos está más sujeto a error que si los datos fuesen continuos y propone que en tal caso sólo habría que usar el análisis ROC para ciertos propósitos. Por último, discute que si la enfermedad a detectar conlleva un tratamiento poco agresivo será mejor subir la sensibilidad a costa de tener falsos positivos y que, sin embargo, si el tratamiento pudiese perjudicar a la salud del paciente habría que subir la especificidad a costa de obtener más falsos negativos.

- Faraggi y Reiser (2001): '*Estimation of the area under the ROC curve*'[18]: Éstos autores hacen una comparativa entre métodos paramétricos y no paramétricos para la construcción de curvas ROC y su posterior cálculo del área. Finalmente aplican dichos resultados a encontrar un prueba que detecte a mujeres portadoras de la enfermedad: distrofia muscular de Duchenne. Para ello cuentan con una muestra compuesta de mujeres portadoras y no portadoras, además de la variable de decisión 'CK' a la cual le aplican diferentes transformaciones obteniendo así diferentes estimadores para el área.

- Martinez-Camblor (2007): '*Comparación de pruebas diagnósticas desde la curva ROC*'[19]: En este artículo se hace un recorrido teórico sobre la construcción de curvas ROC, comparando varias pruebas mediante su área bajo la curva asociada para finalmente aplicarlo a una muestra de pacientes con infección y sin ella para determinar cuál de las tres variables que propone determina mejor su condición.

- Chang et. al (2014): '*Biomarker selection for medical diagnosis using the partial area under the ROC curve*'[69]: Este artículo propone un estudio sobre elección y combinación de variables llamadas 'marcadores' siendo éstas clasificadas en función del área bajo la curva que obtienen para clasificar tres tipos de enfermedades: distrofia muscular de Duchenne, anomalía en el corazón y anomalías en el tejido mamario. En cada uno de los casos aplican el modelo binormal pues suponen la normalidad de todos sus datos.

Otras aplicaciones en medicina

- Metz (1989): 'Some practical issues of experimental design and data analysis in radiological ROC studies'[68].

- Zweig et. al (1993): 'Receiver-Operating Characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine'[14].

- Zou (1998): 'Original smooth receiver characteristic curves estimation from continuous data: statistical methods for analyzing the predictive value of spiral CT of urethral stones'[48].

- Kazmierczak (1999): 'Statistical techniques for evaluating the diagnostic utility of laboratory tests'[67].

- Pepe (2003): 'The statistical evaluation of medical for classification and prediction'[66].

- Xinzhe et. al (2012): 'Three dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value and tumor stage'[65].

- Keunyoung et. al (2012): 'Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small-cell lung cancer'[64].
- Soussan et.al (2014): 'Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer'[63].
- Teixeira (2013):'Avaliação de classificadores na classificação de radiografias de tórax para o diagnóstico de pneumonia infantil'[62].

Psicología

- Navarro et. al (1998): '*El análisis de curvas ROC en estudios epidemiológicos de psicopatologia infantil: aplicación al cuestionario CBCL*'[61]: El objetivo de este artículo es someter al análisis ROC el cuestionario para la detección de problemas psicológicos en niños 'Child Behavior Checklist'. Este consta de un cuestionario donde se evalúa de 1 a 3 la frecuencia de ciertos comportamientos, como resultado se obtienen 10 valores cuantitativos correspondientes a diferentes patologías para cada individuo que, se unifican en un sólo valor representativo de la psicopatología general del individuo. Dicha variable es contrastada tres veces, sobre la misma muestra, con tres criterios diferentes que hacen las veces de variable 'estado'. A saber, si procedían de consulta pediátrica o psiquiátrica (I), el resultado obtenido en una entrevista a los padres (II) y la conclusión de un test que rellenaba cada psicólogo responsable de cada individuo (III).

  Por último, aplican la prueba *t de Student* para la comparación de medias, construyen las curvas ROC con procedimientos no paramétricos, estiman sus áreas y eligen el punto de corte óptimo. En sus resultados obtienen que para los criterios I y II las medias son significativamente distintas, que el área bajo la curva correspondiente al criterio I es mayor a la del II y se desecha la del III pues aunque obtiene un área mayor de 0,5 su intervalo de confianza sí que incluye este valor. Finalmente, para la elección del punto de corte determinan que es preferible una alta sensibilidad y cribar los falsos positivos mediante una segunda prueba. Para ello se quedan con los valores de máxima especificidad que tengan una sensibilidad mayor del 75 %.

Otras aplicaciones en psicología

- Swets et. al (1961): 'Decision processes in perception'[60].
- Erdman et. al (1987): 'Suicide risk prediction by computer interview: a prospective study'[59].
- Olin et. al (1995): 'Assesing the predictive value of teacher reports in high risk sample for schizophrenia: a ROC analysis'[58].

- Swets (1996): 'Signal detection theory and ROC analysis in psychological diagnostic'[57].

Economía

- Frerichs (2003): *'Evaluating internal credit rating systems depending on bank size'*[56]: En este artículo, la curva ROC es una herramienta para clasificar entre buenos y malos prestatarios incentivado por el acuerdo de 2003 en Basilea donde dictaron que cada banco podría tener su propio sistema de clalificación crediticia siempre que éstos cumplan unas especificaciones básicas. Expone que las entidades que se encargan de regular dichos sistema podrían usar las medidas expuestas en el artículo para discriminar entre los que clasifiquen bien y mal. Además del área bajo la curva usa la puntuación de Brier para determinar el valor umbral que mejor capacidad discriminante tiene.

Otras aplicaciones en economía

- Sobehart et. al (2001): 'Measuring default accurately'[55].
- Kennedy et. al (2009):'Low-Default Portfolio/One-Class Classification: A Literature Review'[54].

Meteorología

- Ancell (2013): '*Aportaciones de las redes Bayesianas en meteorología. Predicción probabilística de precipitación*'[53]: El objetivo del autor es predecir sobre áreas locales de la Cornisa Cantábrica la probabilidad de precipitación. Dicha variable es dicotomizada considerando que está presente cuando se recojan más de $2^{mm}/_{24h}$. Describe la curva ROC como la representación de los pares (FAR, HIR), *False Alarm Rate* y *Hit Rate* respectivamente, que equivalen a los valores definidos de ($1-E$, S). Expone que su elección de punto de corte será aquel cuya pendiente de la recta tangente sea 1 pues le supone un resultado válido para maximizar los resultados acertados.

  La toma de datos sobre las predicciones en el territorio de estudio es discretizada a 100 estaciones de la Agencia Estatal Meteorológica, estándo repartidas de forma homogénea por todo el área y recogiendo datos diariamente durante 90 días. Para su análisis diferencia entre *predictando* y *predictores*: lo primero es la variable objetivo, es decir, ¿habrá o no habrá precipitación? y lo segundo son aquellas variables que ayudan a predecir el estado del predictando (temperatura, dirección, viento, humedad...). Debido al coste computacional que tendría llevar el análisis a cabo con 100 observaciones n-dimensionales diarias a lo largo de 90 días aplica un análisis de componentes principales. Dicho análisis determina que agrupe sus variables explicativas en dos componentes para los cuales calcula sus

centroides. Por último, aplicando redes bayesianas obtiene la estimacion de la curva ROC con un área de aproximadamente del 80 %.
El trabajo de esta tesis continúa ampliando el terreno a toda la Península Ibérica e Islas Baleares, considerando diferentes valores umbrales, aumentando el tiempo de estudio a 45 años y, además, aplicando diferentes redes con diferentes restricciones a la hora de conectar los predictores.

Otras aplicaciones en meteorología

- Harvey et. al (1992): 'Application of signal detection theory to weather forecasting behaviour'[52].
- Wilks (2001): 'A skill score based on economic value for probability forecast'[51].

Aprendizaje automático

- Davis y Goadrich (2006): *'The relationship between Precision-Recall and ROC curves'*[50]: Con el objetivo de comparar la eficiencia de dos algoritmos con resultados dicotómicos, éstos autores construyen sus correspondientes curvas ROC y Precision-Recall, alegan que un simple índice de exactitud no es suficiente y que a menudo, en aprendizaje automático, surje la necesidad de validar un nuevo algoritmo. Sostienen que cuando este tiene respuestas binarias ha de contrastarse mediante un análisis ROC. Las curvas precision-recall (PR) se obtiene al representar los pares:

$$(\frac{V_+}{V_+ + F_-}, \frac{V_+}{V_+ + F_+})$$

Nótese que es igual a (FVP, VPP), es decir, sensibilidad frente al valor predictivo positivo respectivamente. Además, demuestran que dada una curva ROC exite una única curva PR equivalente a ella y, de hecho, tengan la misma tabla de contingencia. Por otro lado, la curva ROC tiene más capacidad discriminante cuanto más se aproxime al $(0, 1)$ y, sin embargo, la PR cuanto más se aproxime al $(1, 1)$ pero ocurre que si dadas dos curvas ROC una discrima mejor que otra entonces sus correspondientes curvas PR guardan la misma relación. También se da un procedimiento para interpolar puntos de una curva con el objetivo de hacer perfectamente cóncava su curvatura y se demuestra que los puntos a excluir en la curva ROC (por estropear la concavidad) son los mismos que se escluyen en la PR, no en coordenadas sobre el cuadrado $[0, 1] \times [0, 1]$ pero si en equivalencia, es decir, los calculados bajo el mismo valor umbral.

Otras aplicaciones en aprendizaje automático

- Spackman (1989): 'Signal detection theory: valuable tools for evaluating inductive learning'[21].
- Prati et. al (2008):'Curvas ROC para avaliação de classificadores'[20].

Conclusión

En este trabajo, tras analizar y presentar los conceptos de *sensibilidad* y *especificidad*, hemos recogido un estudio teórico de la construcción de la curva ROC, acompañado de sus diferentes métodos, a saber: paramétricos, no paramétricos y semiparamétricos. El paramétrico presenta el problema de confirmar mediante contraste de hipótesis que los resultados de la variable de decisión corresponden a la distribución escogida, el no paramétrico puede representarnos una curva ROC muy escalonada si el tamaño de la muestra es pequeño y el semiparamétrico supone la existencia, no asegurada, de una transformación monótona que aproxime los datos a una distribución normal.

En el Capítulo 3 hemos visto varias formas de cuantificar la capacidad discriminante de un clasificador. El término *accuracy* representaba la probabilidad *a posteriori*, es decir, una vez hecha la clasificación qué porcentaje es el correcto, el *índice de Youden* nos permite encontrar el punto de corte de mayor sensibilidad y especificidad conjunta, la *tasa de verosimilitud* nos ha proporcionado una medida de exactitud de clasificación de los individuos con presencia del evento y otra para los individuos con ausencia, el *índice de discriminación* mide el solapamiento de las funciones de densidades de sanos y enfermos, es decir, es una medida de cuánto le cuesta a la variable de decisión diferenciarlos y el *odds ratio* definide una medida de ocurrencia de respuesta positiva en presencia y ausencia del evento. Sin embargo, la medida más usada por autores de diferentes campos es el *área bajo la curva*, esta no sólo nos aporta una medida de la bondad del clasificador sino que nos permite comparar pruebas y, en caso de igualdad, compararlas mediante las áreas parciales.

Para la elección del *punto de corte* podemos hacer uso del índice de Youden, sin embargo, este presenta el problema de maximizar la sensibilidad y especificidad conjuntamente. Una alternativa a este problema es escoger el punto de corte correspondiente al par $(1-E, S)$ más cercano al vértice $(0, 1)$ (Lema 1, Capítulo 4), o bien, el que minimice los costes de resultados erróneos (Lema 2, Capítulo 4), ambos métodos dependen de la prevalencia del evento.

Las anteriores definiciones se basan en los conceptos de sensibilidad y especificidad, estas medidas teórica, poblacionales, pueden estimarse mediante la *fracción de verdaderos positivos* y la *fracción de verdaderos negativos* respec-

tivamente, ambas cantidades muestrales. Por tanto, el análisis ROC será más preciso cuanto mejor represente la muestra seleccionada a la población.

En el Capítulo 5 hemos construido las curvas ROC correspondientes a los datos de dos biomarcadores medidos a individuos con y sin cáncer de páncreas. Los programas usados han sido SPSS y R obteniendo en ambos la misma conclusión respecto a qué biomarcador discrimina mejor la muestra. La diferencia entre ambos es que SPSS sólo permite hacer estimaciones paramétricas de la curva ROC usando la distribución exponencial binegativa, mientras que, las funciones definidas en el paquete pROC de R permiten suavizar la curva mediante diversas distribuciones además de calcular áreas parciales, intervalos de confianza para la sensibilidad y especificidad y hacer un contraste de hipótesis de igualdad de pruebas.

Fuera del campo de las Matemáticas, las curvas ROC son usadas en diversas áreas como hemos visto en el Capítulo 6, en especial en Medicina, cuando ante una enfermedad intentan encontrar el marcador o la medida que la detecte sin necesidad de un *test de oro (golden test)*, es decir, una prueba perfecta sin error en sus resultados, la correspondiente a nuestra variable $D = estado$, ya sea para detección precoz de enfermedades o porque dicha prueba suponga un peligro para la salud del paciente.

Finalmente, las curvas ROC suponen una eficiente herramienta para medir la bondad de un clasificador y, en caso de ser suficiente, elegir el valor umbral que mejor se ajuste al evento a detectar. Es decir, una prueba más sensible que específica (preventiva), una más específica que sensible (tamizaje) o una que minimice el o los resultados erróneos positivos y negativos. Puede realizarse para todo tipo de dato (continuo, discreto, categórico...), la sustenta una base teórica que cubre métodos de estimación, cuantificadores de su exactitud y contrastes de hipótesis para aceptar o no la igualdad de pruebas y salvar el problema de la variabilidad muestral.

Bibliografía

[1] Green DM, Swets JA(1966): 'Signal detection theory and psichophysics'.New York: John Wiley & Sons, Inc.

[2] Lusted LB (1971): 'Decision making studies in patient management'. N Engl J Med; 171:1217-9.

[3] Lusted LB (1971): 'Signal detectability and medical decision-making'. Science; 171:1217-9.

[4] Ataque a Pearl Harbor: Enlace: https://es.wikipedia.org/wiki/Ataque_a_Pearl_Harbor.

[5] Burgueñoa M.L, García-Bastos J.L y González-Buitrago J.M. (1995) :'Las curvas ROC en la evaluación de las pruebas diagnósticas'.Med Clin (Barc).104: 661-670

[6] Torres A. (2010): 'Curvas ROC para Datos de Supervivencia. Aplicación a Datos Biomédicos'. Trabajo final de master en Técnicas Estadísticas de la Universidad de Santiago de Compostela.

[7] López de Ullibarri G. I, Píta Fernández S. (1998): 'Curvas ROC'. Cad Aten Primaria. 5: 229-235.

[8] Hanley JA (1988): 'The robustness of the binormal model used to fit ROC curves'. Med Decision Making 8:197-203.

[9] Swets JA.(1988): 'Measuring the accuracy of diagnostic systems'. Science 240: 1.285-1.293.

[10] Bamber D. (1975):'The area above the ordinal dominance graph and the area below the receiver operating graph'. J Math Psych; 12: 387-415.

[11] Hanley JA, McNeil BJ (1983): 'A method of comparing the areas under receiver operating characteristic curves derived from the same cases'. Radiology; 148: 839-843.

[12] Armesto D. (2011): 'Pruebas diagnósticas: curvas ROC'.Electron J Biomed;1:77-82.

[13] Schisterman EF, Perkins NJ, Liu A, Bondell H (2005): 'Optimal Cut-point and Its Corresponding Youden Index to Discriminate Individuals Using Pooled Blood Samples'. Epidemiology; 16(1): 73-81.

[14] Zweig MH, Campbell G (1993): 'Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine'. Clin Chem; 39: 561-577.

[15] McNeil BJ, Keeler E, Adelstein SJ (1975): 'Primer on certain elements of medical decision making'. N Engl J Med; 293: 211-215.

[16] Franco N y Vivo Molina J.M, 'Análisis de curvas ROC: Principios básicos y aplicaciones'. Cuadernos de Estadística. Ed: La Muralla, S.A. ISBN: 978-84-7133-772-6.

[17] García J.J, 'Apuntes sobre razones de probabilidad o verosimilitud'. Universidad Nacional Autónoma de México.

[18] Faraggi D, Reiser B (2002): 'Estimation under the ROC curve'. Statist. Med. 21:3093-3106.

[19] Martínez-Camblor P (2007): 'Comparación de pruebas diagnósticas desde la curva ROC'. Revista Colombiana de Estadística. 2:163-176.

[20] Prati R.C, Batista G.E, Monard M.C.(2008):'Curvas ROC para avaliação de classificadores'. IEEE latin america transactions. Vol 6, Nº2.

[21] Spackman (1989): 'Signal detection theory: valuable tools for evaluating inductive learning'.Morgan Kaufmann Publishers Inc. 1-55860-036-1.

[22] 'Función Kernel'. Ecured. Enlace: https://www.ecured.cu/Funci%C3%B3n_Kernel

[23] Metz CE, Herman BA, Shen J. (1998): 'Maximum likehood estimation of reciever operating characteristic (ROC) curves from continuously distributed data. Statisc Med; 17:1033-1053.

[24] Gonzalves L, Subtil A, Oliveira M.R y Bermudez P.Z (2014): 'ROC Curves Estimation: An overview'. Revstat-Statiscal Journal, Volume 12, Number 1.

[25] John Eng, M.D, 'ROC Analysis. Web-based Calculator for ROC Curves', Johns Hopkins University School of Medicine. Enlace: http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html

[26] Lusted, L.B (1971): 'Signal detectability and medica decision-making'. Science, 171, 1271-9.

[27] 'Diagnostic and Biomarkers Statistical (DABS) Center'/Datasets, enlace: http://research.fhcrc.org/diagnostic-biomarkers-center/en/datasets.html

[28] Robin X et al. (2011). 'pROC: an open-source package for R and S+ to analyze and compare ROC curves'.

[29] Youden, W.J. (1950). "Index for rating diagnostic tests". Cancer. 3: 32–35

[30] Zhou X.H, Hall W.J, Shapiro, D.E (1997): 'Smooth non-parametric reciever operatin characteristic (ROC) curves for continuous diagnostic tests'. Statist. Med. 16, 2143-2156.

[31] Lloyd C.J (1998): 'The use of smoothed ROC curves to summarise and compare diagnostic systems'. J. Amer. Statisc. Assoc. 93, 1356-1364.

[32] Zweig M.H y Campbell G (1993):'Reciever-Operating Characteristic (ROC) Plots: A fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561-577.

[33] Dorfman D.D y Alf E (1969): 'Maximum likehood estimation of parameters of signal-detection theory and determination of confidence interval-rating method data'. J. Math. Psychol. 6, 487-496.

[34] Turnbull B.W, Hsieh F (1996): 'Nonparametric and semiparametric estimation of the receiver operating characteristic curve'. The annals of statistics 24, 25-40.

[35] Farcomeni, A. y Ventura, L.(2010): 'An overview of robust methods in medical research'. Statistical Methods in Medical 21.

[36] Heritier, S. Cantoni, E. Copt, S. y Victoria-Feser, M.-P. (2009): 'Robust Methods in Biostatistics'. John Wiley & Sons, Chichester.

[37] Lloyd C.J y Yong Z (1999): 'Kernel estimators of the ROC curves are better than empirical'. Statist. Probab. Letters, 44, 221-228.

[38] Zhou XH1, Harezlak J. (2002): 'Comparison of bandwidth selection methods for kernel smoothing of ROC curves'. 14:2045-55.

[39] Hall P, Hyndman R.J, Fan Y. (2003): 'Nonparametric confidence intervals for receiver operating characteristic curves'.3: 743-750.

[40] Jokiel-Rokita, A. y Pulit, M.S. (2013): 'Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions'. Stat Comput 23: 703.

[41] Pepe M.S y Cai T. (2002): 'The analysis of placement values for evaluating discriminatory measures'. Technical report, University of Washington.

[42] Franco, M, y Vivo, J.M. (2007): 'Análisis de curvas ROC: principios básicos y aplicaciones'. Madrid: La Muralla.

[43] Metz CE y Kronman HB. (1980): 'Statistical significance tests for binormal ROC curves'. Journal of Mathematical Psychology 22: 218-243.

[44] Hanley J.A y McNeil M.D (1982): 'The meaning and use of the area under a receiver operating characteristic (ROC) curve'. Radiology 143:29-36.

[45] Hilden J. (1991): 'The area under the ROC curve and its competitors'. Med Decis Making.11:95-101.

[46] Swets J.A y Pickett R.M (1982): 'Evaluation of Diagnostic Systems'. Academic Press, Inc. New York.

[47] Sox H, et al. (1989): ' Assessment of diagnostic technology in health care. Rationale, methods, problems and directions'. National Academy Press.

[48] Zou K.H, Tempany C.M, Fielding J.R, Silverman S.G (1998): 'Original smooth reciever operating characteristic curves estimation from continuous data: statistical methods for analyzing the predictive value of spiral CT of ureteral stones. Academic Radiology. 5:680-687.

[49] Jaeschke R, Guyatt G, Lijmer J (2002): 'Diagnostic tests'. AMA Press, Chicago, 121-40.

[50] Davis J y Goadrich M (2006): 'The relationship between Precision-Recall and ROC curves'. ACM New York. 1-59593-383-2.

[51] Wilks D.S (2001): 'A skill score based on economic value for probability forecast'.Meteorol. Appl. 8, 209–219.

[52] Lewis O. Harvey Jr., Kenneth R. Hammond, Cynthia M. Lusk y Ernest F. Mross (1992): 'Application of signal detection theory to weather forecasting behaviour'. American Meteorological Society.

[53] Ancell R (2013): 'Aportaciones de las redes Bayesianas en meteorología. Predicción probabilística de precipitación'. Tesis doctoral.

[54] Kennedy K, Mac Namee B, Delany S.J (2009):'Low-Default Portfolio/One-Class Classification: A Literature Review'. Technical Report: SOC-AIG-001-09.

[55] Sobehart J, Keenan S (2001): 'Measuring default accurately'.

[56] Frerichs H. Wahrenburg M (2003): 'Evaluating internal credit rating systems depending on bank size'. Working Paper Series: Finance & Accounting, No. 115.

[57] Swets J.A (1996): 'Signal detection theory and ROC analysis in psychological diagnostic'. Lawrence Erlbaum Associates, 1996 - 308.

[58] Olin S.S, Richard S. John, Sarnoff A. Mednick (1995): 'Assesing the predictive value of teacher reports in high risk sample for schizophrenia: a ROC analysis'. Elsevier Inc. 16. 53–66.

[59] Erdman HP, Greist JH, Gustafson DH, Taves JE, Klein MH. Suicide risk prediction by computer interview: a prospective study. J Clin Psychiatry. 1987;48(12):464–467.

[60] Swets J.A, Birdsall T.G, Tanner W.P (1961): 'Decision Processes In Perception'. Psychological Review 68(5):301-40.

[61] Navarro J.B, Domenech J.M, de la Osa N (1998): 'El análisis de curvas ROC en estudios epidemiológicos de psicopatologia infantil: aplicación al cuestionario CBCL'. Anuario de psicología. 29.3-15.

[62] Teixeira R (2013) :'Avaliação de classificadores na classificação de radiografias de tórax para o diagnóstico de pneumonia infantil'. Dissertação (Mestrado em Ciência da Computação).

[63] Soussan M, Orlhac F, Boubaya M, Zelek L, Ziol M, Eder V, Buvat I (2014): 'Relationship between Tumor Heterogeneity Measured on FDG-PET/CT and Pathological Prognostic Factors in Invasive Breast Cancer'. PLoS ONE 9(4): e94017.

[64] Kim K1, Kim SJ, Kim IJ, Kim YS, Pak K, Kim H. (2012): 'Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small-cell lung cancer'. Nucl Med Commun.33(6):613-20.

[65] Dong X (2012): 'Three dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value and tumor stage'. Nucl Med Commun 34 (1), 40-46.

[66] Pepe M.S (2003): 'The statistical evaluation of medical for classification and prediction'. Oxford Statistical Science Series.

[67] Kazmierczak S.C (1999): 'Statistical techniques for evaluating the diagnostic utility of laboratory tests'. Clin Chem Lab Med. 37(11-12):1001-9.

[68] Metz C.E (1989): 'Some practical issues of experimental design and data analysis in radiological ROC studies'. Invest Radiol.(3):234-45.

[69] Chang Y. C, Hsu M.J, Hsueh H.M (2014): 'Biomarker selection for medical diagnosis using the partial area under the ROC curve'. BMC Res Notes. 10.1186/1756-0500-7-25.

[70] Erkel A.R y Pattynama P.M (1998): 'Receiver-Operating Characteristic (ROC) analysis: basic principles and applications in radiology. Eur J Radiol.(2):88-94.