

# Curvas ROC

*00R Team*

*Abril 2016*

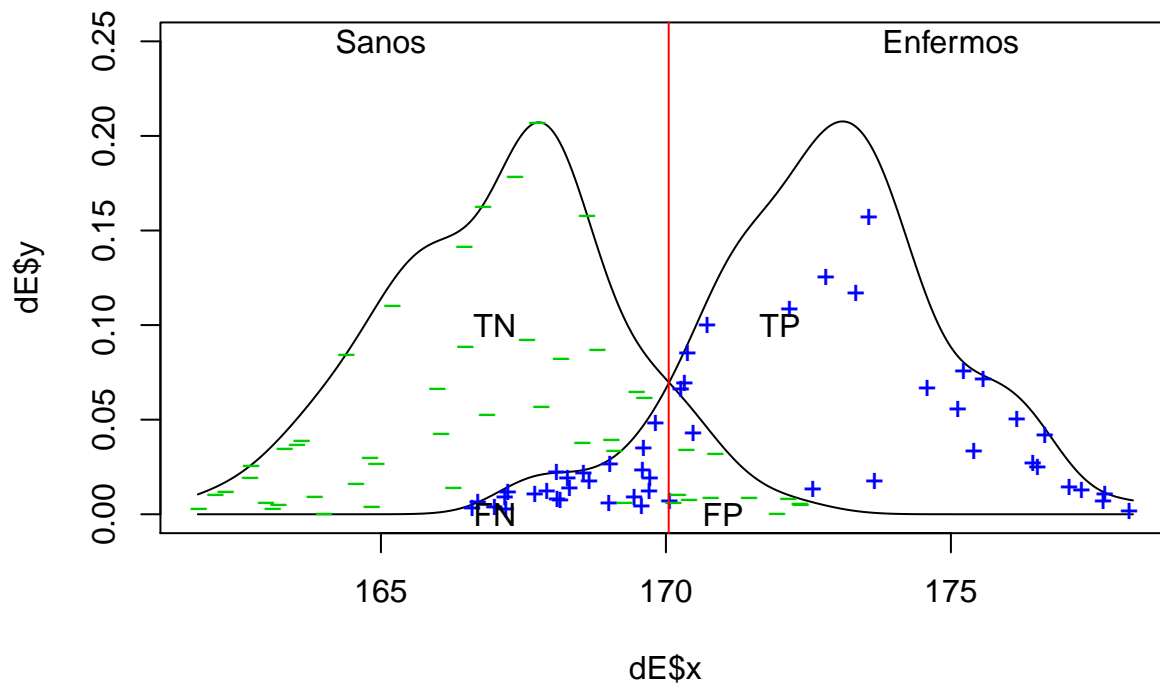
## Índice

1. Qué son y para que sirven las curvas ROC	1
2. Curvas ROC más sencillas y vistosas	6
2.1. ROCR . . . . .	6
2.2. pROC . . . . .	6
3. Punto de corte y área bajo la curva	7
4. Ejercicio	7

## 1. Qué son y para que sirven las curvas ROC

- Una curva ROC (*Receiver Operating Characteristic*) es una representación gráfica que ilustra la relación entre la sensibilidad y la especificidad de un sistema clasificador para diferentes puntos de corte.
- Fue desarrollada por ingenieros eléctricos durante la II Guerra Mundial, para medir la eficacia de la detección de objetos enemigos en el campo de batalla mediante señales de radar.
- Su uso está muy extendido en medicina, para validar técnicas diagnósticas. Más recientemente con el auge de las técnicas de aprendizaje automatizado, se han empleado las curvas ROC para evaluar diferentes algoritmos de clasificación.

Crearemos un ejemplo para entender el funcionamiento y la aplicación, una población donde una prueba proporciona resultados con distribución normal, con una media para los enfermos de 173 y para los sanos 167 y, en ambos casos la, desviación típica es de 2.



## [1] 0.52

## [1] 0.78

La sensibilidad (ratio de verdaderos positivos (*True Positive*,  $TP$ ) sobre el total de positivos) es importante cuando la enfermedad no puede pasar desapercibida. Es decir, cuando es más grave diagnosticar a un enfermo como sano que diagnosticar a un sano como enfermo.

La especificidad (ratio de verdaderos negativos (*True Negative*,  $NP$ ) entre todos los negativos) es importante cuando la enfermedad puede suponer una estigmatización del paciente o no se quiere preocupar a un paciente o el tratamiento puede ser lesivo (amputaciones). Es decir cuando es más “grave” diagnosticar a un sano como enfermo que pasar desapercibido a uno.

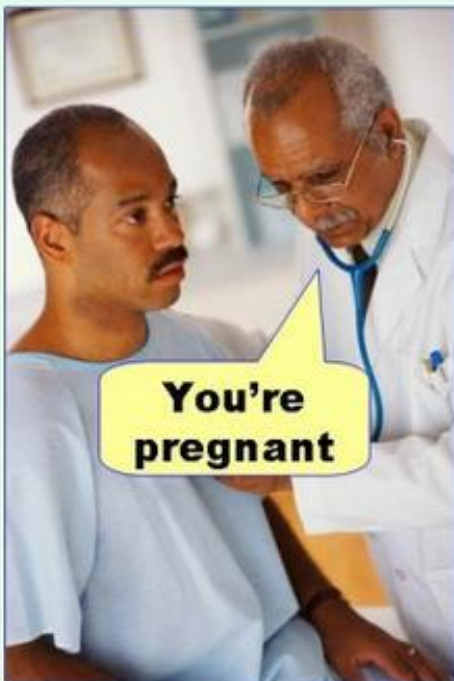
$$Sensibilidad = \frac{TP}{FN + TP}$$

$$Especificidad = \frac{TN}{FP + TN}$$

Lo que interesa es aumentar la sensibilidad y la especificidad. Pero ¿es posible?

¿Qué pasa si intentamos aumentar la sensibilidad? Para ello necesitamos aumentar el número de TP (verdaderos positivos), y eso lo conseguimos desplazando el punto de corte a la izquierda.

**Type I error**  
(false positive)



**Type II error**  
(false negative)

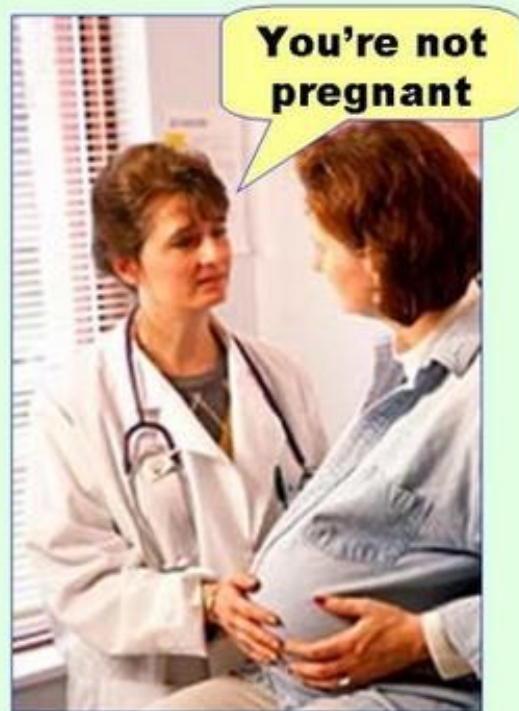
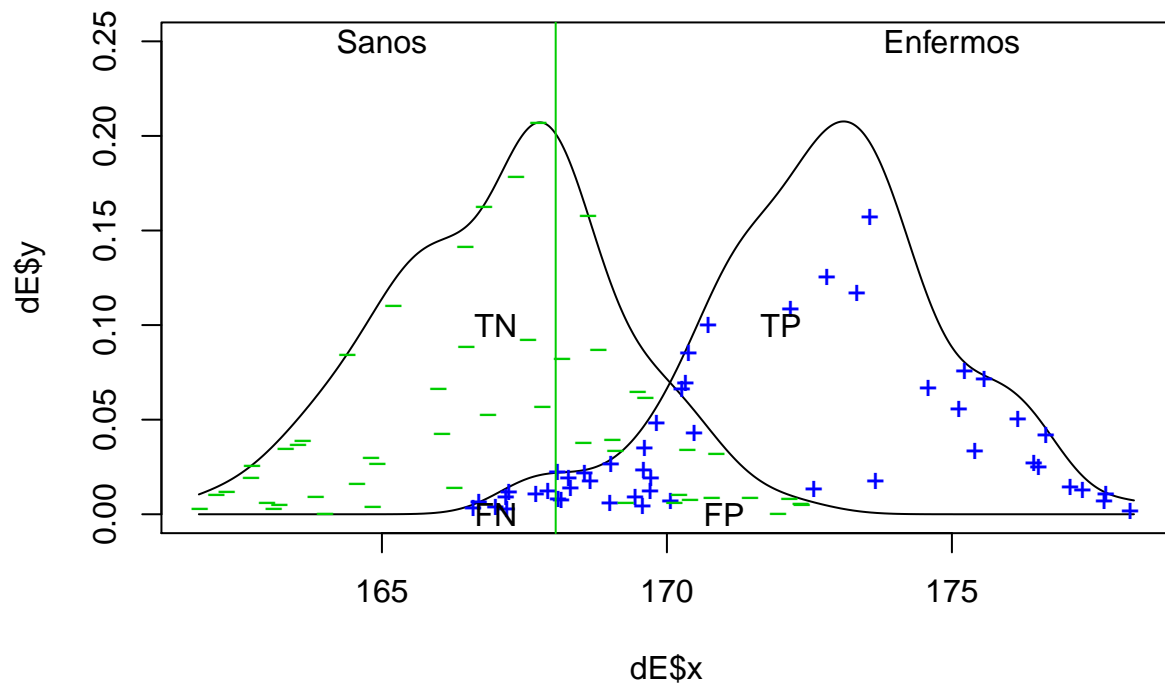
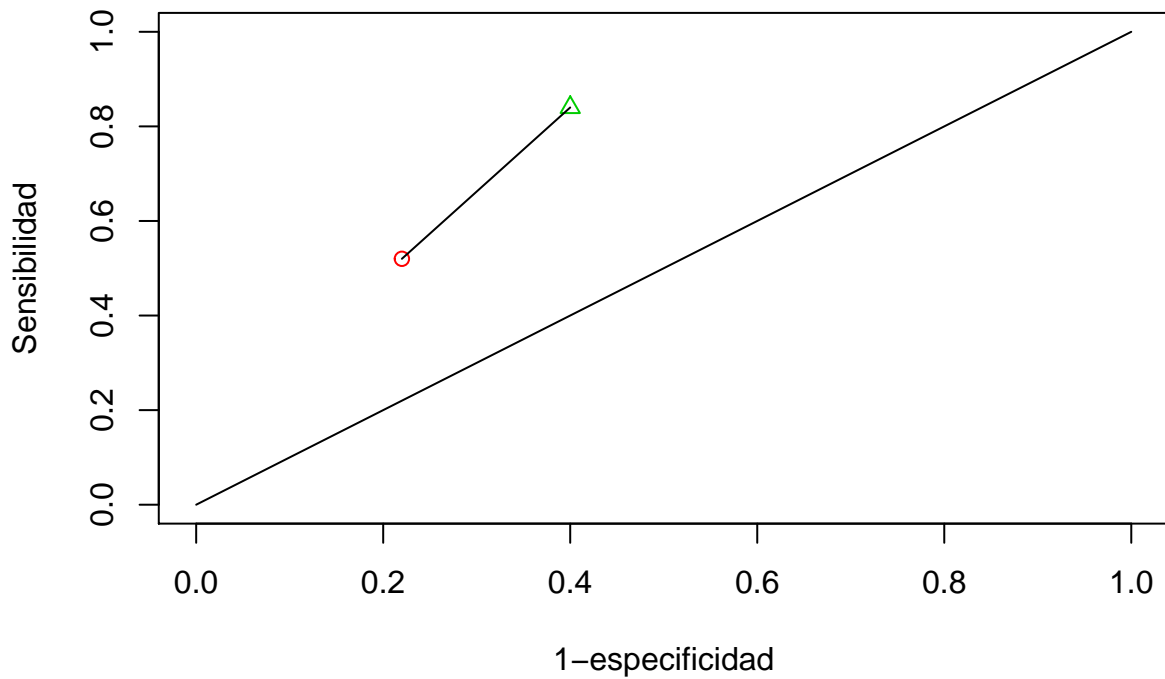


Figura 1:

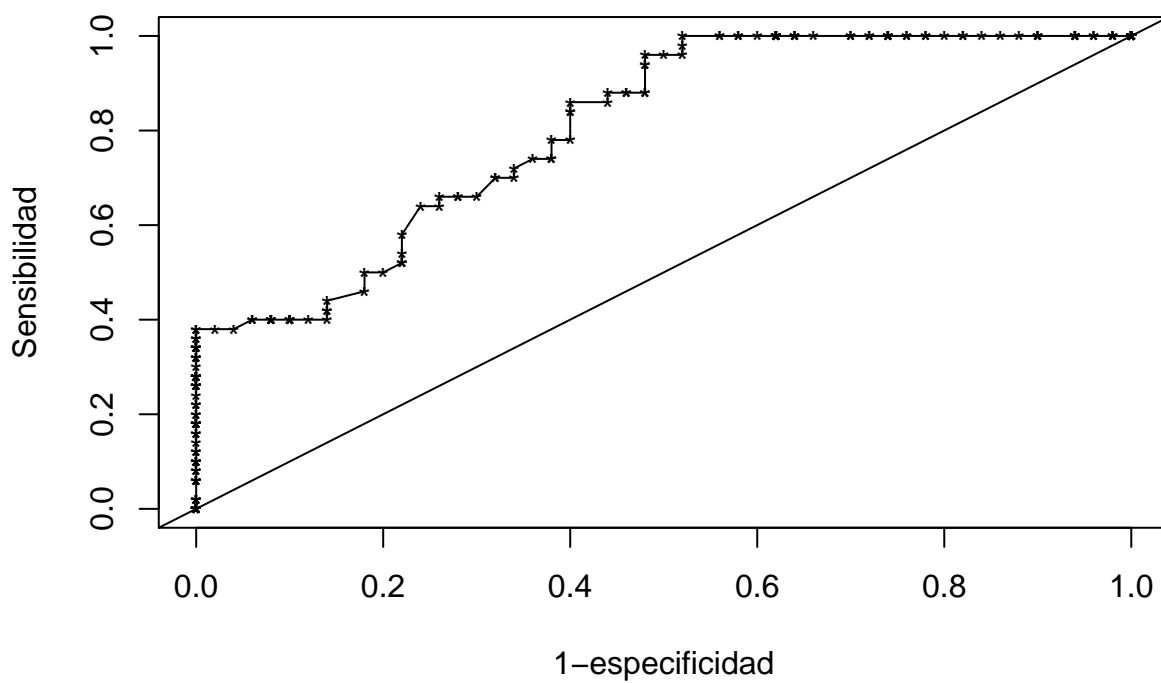
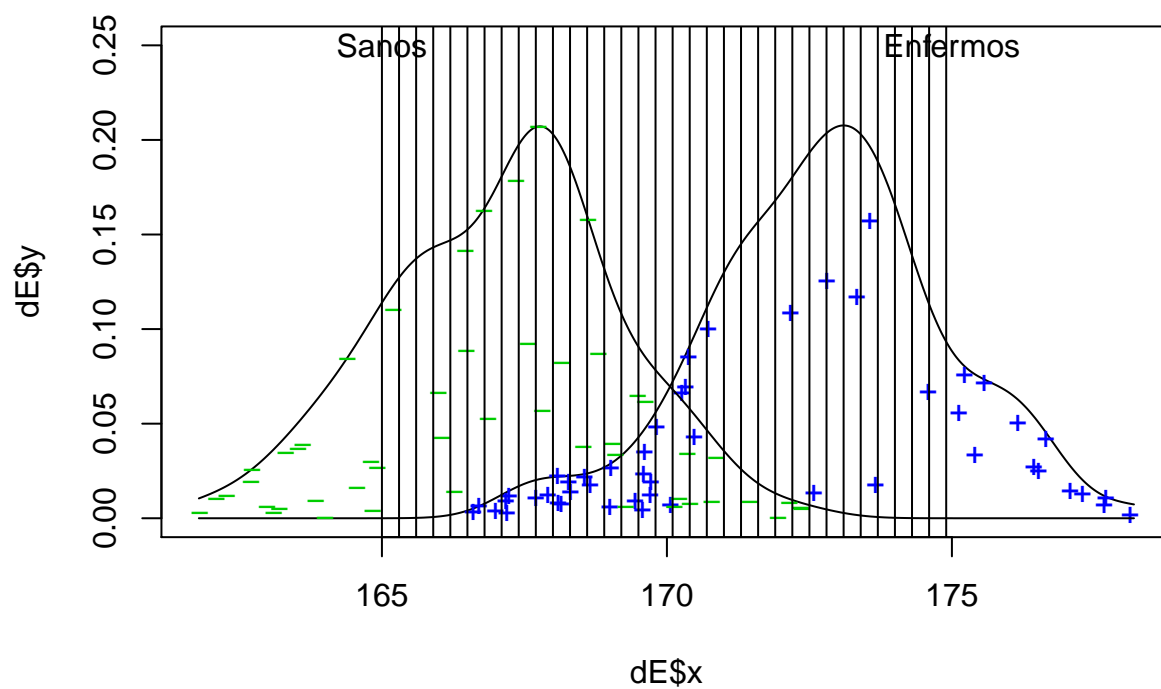


## [1] 0.84

## [1] 0.6

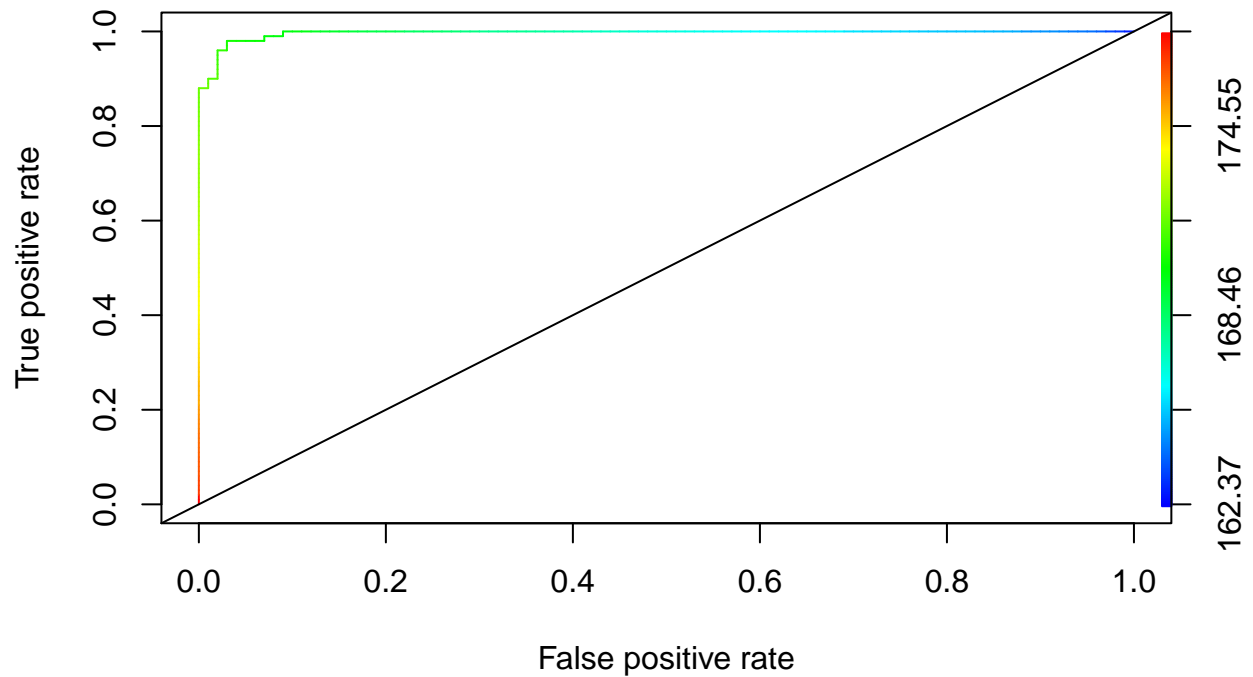


Si esto se hiciese para un rango grande de puntos de corte:



## 2. Curvas ROC más sencillas y vistosas

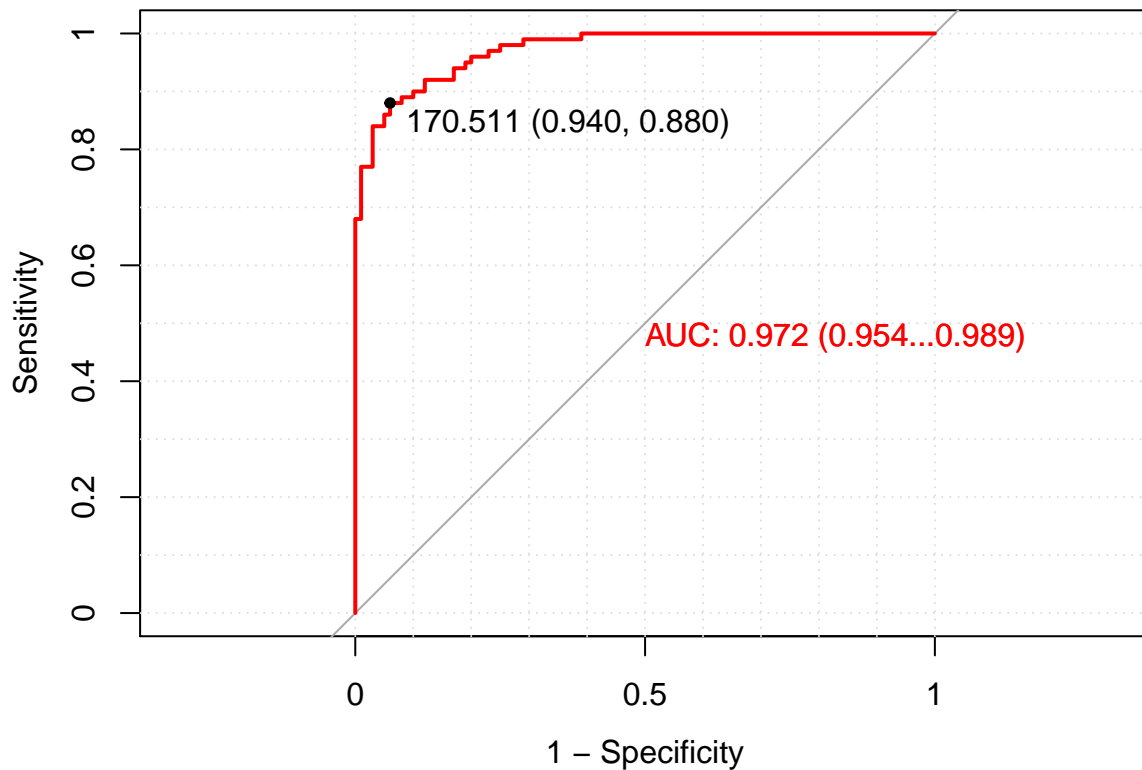
### 2.1. ROCR



```
## AUC: 0.9964
```

### 2.2. pROC

```
##  
## Call:  
## roc.default(response = datos$attr, predictor = datos$val, auc = TRUE,      ci = TRUE)  
##  
## Data: datos$val in 100 controls (datos$attr F) < 100 cases (datos$attr T).  
## Area under the curve: 0.9717  
## 95% CI: 0.9542-0.9892 (DeLong)
```



### 3. Punto de corte y área bajo la curva

Aplicación shiny ROC

El punto de corte óptimo dependerá de lo que se está buscando. Si lo que queremos es maximizar sensibilidad y especificidad se puede utilizar el índice de Youden, donde para cada par sensibilidad-especificidad, el punto de corte sería aquel donde se cumpla:

$$\text{Max}( \text{sensibilidad} + \text{especificidad} - 1 )$$

```
## [1] 170.5115
```

### 4. Ejercicio

- El fichero IMC.dat contiene datos con el índice de masa corporal (IMC) de 200 individuos y si están clasificados con sobrepeso (T) o sin sobrepeso (F).
- Por otro lado el fichero peso.dat contiene el peso de 200 individuos igualmente clasificados con sobrepeso y sin él.
- Se quiere determinar cuál de las 2 “pruebas diagnósticas” sería la más adecuada para determinar si un individuo tiene sobrepeso y cuál sería el punto de corte óptimo.
- Representar ambas curvas en el mismo gráfico (ver las ayudas de las funciones).