

Curvas ROC y Análisis de Supervivencia.

Aplicación a datos de pacientes ingresados por Infarto Agudo de Miocardio

Introducción

La curva ROC (*Receiver Operating Characteristic*) desempeña un papel fundamental en la evaluación de la precisión de las pruebas usadas para diagnosticar enfermedades. Cuando el objetivo de la prueba diagnóstica es la predicción de eventos en el tiempo, se involucran métodos de supervivencia y variables tiempo dependientes. En este contexto, la evaluación de la precisión de las pruebas diagnósticas requiere extender los conceptos tradicionales de sensibilidad y especificidad. En los últimos años se han propuesto varias definiciones y métodos de estimación de curvas ROC tiempo-dependientes. El propósito de este trabajo es realizar una revisión de estas definiciones, e ilustrar su uso a través de su aplicación a datos de pacientes ingresados por Infarto Agudo de Miocardio.

Curva ROC

Sea Y un marcador continuo. La clasificación de un sujeto como enfermo ($D = 1$) o sano ($D = 0$) puede realizarse mediante la dicotomización de Y , eligiendo un valor de corte c y definiendo el resultado de la prueba como positivo para enfermedad si $Y \geq c$ (es decir, el individuo es clasificado como enfermo). Se define la curva ROC como el gráfico de la *fracción de verdaderos positivos* (TPF ó Sensibilidad) versus la *fracción de falsos positivos* (FPF ó 1 - Especificidad) para todos los posibles valores de corte c :

$$ROC(\cdot) = \{FPF(c), TPF(c), c \in (-\infty, \infty)\},$$

donde $TPF(c) = P(Y \geq c | D = 1)$

$$FPF(c) = P(Y \geq c | D = 0)$$

Curva ROC Tiempo Dependiente

En el caso de estudios de supervivencia, la definición tradicional de la variable D (enfermo vs sano) cambia, pues es necesario incorporar el tiempo T de supervivencia. Por ejemplo, en este contexto puede ser de interés clasificar, a partir del marcador Y , a los individuos que presentan un evento (p.e. la muerte) antes de un tiempo t determinado, de aquellos que lo presentan a posteriori. En esta situación, se define una nueva variable binaria, $D(t)$, que representa el "status" de un individuo en un tiempo t :

$$D(t) = \begin{cases} 1 & \text{si } T \leq t \\ 0 & \text{si } T > t \end{cases}$$

Diferentes definiciones de curva ROC tiempo dependiente se han propuesto en la literatura estadística, que difieren en la definición de la variable $D(t)$:

1. ROC Acumulativa-Dinámica (Heagerty et al.(2000))

$$FPF_t^D(c) = P(Y \geq c | T > t)$$

$$TPF_t^A(c) = P(Y \geq c | T \leq t)$$

2. ROC Incidente-Dinámica (Heagerty and Zheng (2000))

$$FPF_t^D(c) = P(Y \geq c | T > t)$$

$$TPF_t^I(c) = P(Y \geq c | T = t)$$

3. ROC Incidente-Estática (Heagerty and Zheng (2000))

$$FPF_{t^*}^E(c) = P(Y \geq c | T > t^*)$$

$$TPF_t^E(c) = P(Y \geq c | T = t)$$

Referencias

- [1] P. Heagerty, T. Lumley, M. Pepe.(2000). Time Dependent ROC Curve for Censored Survival Data and a Diagnostic Marker *Biometrics*,56(2):337-344
- [2] P. Heagerty, Y. Zheng.(2005). Survival Model Predictive Accuracy and ROC Curves *Biometrics*,61:92-105

Aplicación a datos de Infarto Agudo de Miocardio

Descripción de la muestra

El conjunto de datos está constituido por 811 observaciones (registradas en el periodo de Septiembre del 2003 a Marzo del 2007) correspondientes a individuos que ingresaron con Infarto Agudo de Miocardio en el Servicio de Cardiología del Complejo Hospitalario Universitario de Santiago. Del total de individuos, 253 son diabéticos.

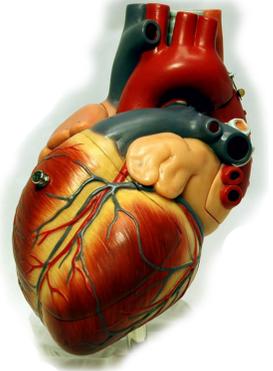
El evento final de interés es la muerte del sujeto. El 81% del total de tiempos de vida son censurados.

En este trabajo se ha considerado como biomarcador el *score* del modelo multivariante de Cox flexible (usando natural splines) que incorpora las siguientes variables: edad, mdrd (fallo renal), glucosa en ayunas, cateterismo, anemia, hipertensión arterial, sexo, sca, cirugía previa, status de fumador.

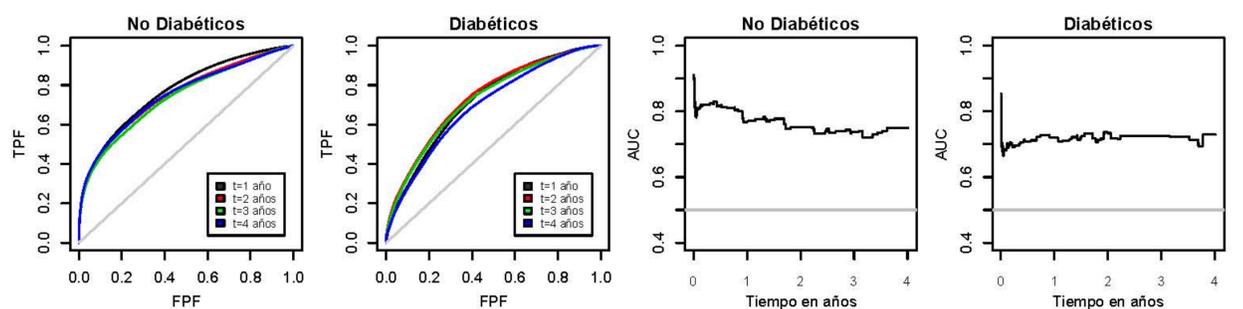
Todos los análisis se han realizado separadamente en los individuos diabéticos y no diabéticos.

Resultados

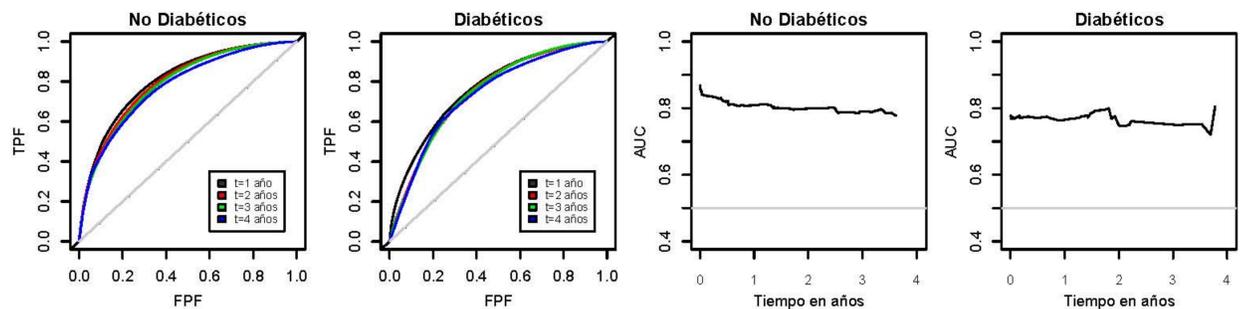
Utilizado los paquetes *survivalROC* y *riksetROC* del software R, se han estimado las curvas ROC tiempo dependiente (1-3). Se han obtenido también las áreas bajo la curva (AUC, *Area Under the Curve*) para tener un resumen general del comportamiento del biomarcador.



(a) Curva ROC y AUC Acumulativa-Dinámica



(b) Curva ROC y AUC Incidente-Dinámica



(c) Curva ROC y AUC Incidente-Estática

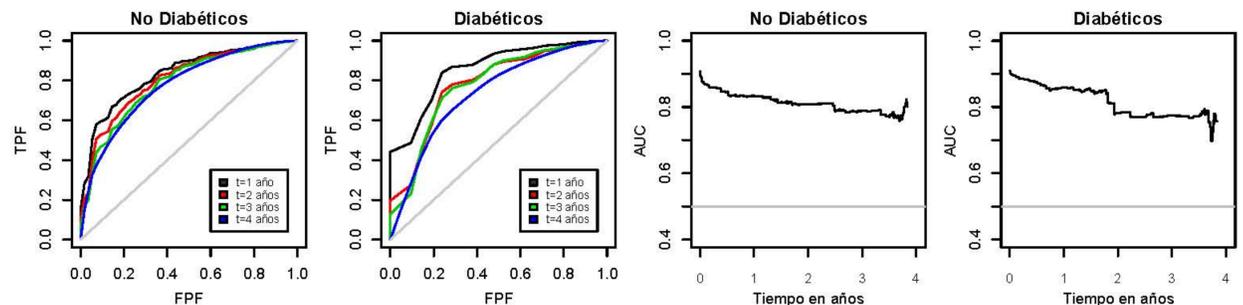


Figura 1: A la izquierda, curvas ROC generadas para no diabéticos y diabéticos. Cada gráfica muestra 4 curvas en colores distintos correspondientes a 4 tiempos de interés (1, 2, 3 y 4 años) en el estudio clínico de supervivencia de los pacientes con Infarto Agudo de Miocardio. A la derecha AUC para no diabéticos y diabéticos, considerando todos los tiempos de interés. Para generar todas las gráficas se utilizó: (a) la definición de curva ROC Acumulativa-Dinámica, (b) la definición de curva ROC Incidente-Dinámica; y, (c) la definición de Curva ROC Incidente-Estática.

Conclusiones Generales

(1) En general, las AUC para todos los tiempos considerados es superior a 0.7, utilizando cualquier definición de curva ROC tiempo dependiente. Estos resultados muestran que el biomarcador construido a partir de las variables consideradas presenta un comportamiento satisfactorio, revelando el interés de aplicar esta prueba diagnóstica (biomarcador) a pacientes con Infarto Agudo de Miocardio. (2) También de forma general, las AUC para los individuos diabéticos son ligeramente inferiores a las AUC de los individuos no diabéticos. (3) Con el uso de estas curvas ROC tiempo dependientes se puede dar respuestas a diversas preguntas clínicas sobre la supervivencia de sujetos, que involucran varios tiempos de interés.

Agradecimientos

Este trabajo fue financiado parcialmente por el Ministerio de Ciencia e Innovación bajo el proyecto: MTM2008-01603.