Modelos de regresión de Cox para el tiempo de supervivencia

Preparado por Luis M. Molinero (Alce Ingeniería)
CorreoE: **bioestadistica@alceingenieria.net**

Artículo en formato PDF

Julio 2001

Modelos de regresión para el tiempo de supervivencia

Tal y como se ha descrito en el artículo anterior, en el análisis de supervivencia la variable de interés es el tiempo hasta que ocurre un suceso. Un posible método para este análisis consiste en suponer que esos tiempos siguen una determinada distribución o función matemática. Para ello se plantea un modelo de cómo evoluciona en función del tiempo la **tasa de mortalidad**. Una vez determinada la tasa de mortalidad, se calcula, a partir de ella, la función de supervivencia.

La tasa de mortalidad también se denomina **función de riesgo** y se suele representar como $\lambda(t)$. El caso más sencillo corresponde a considerar que no varía a lo largo del tiempo, sino que es constante. Se puede comprobar entonces que la función de supervivencia es:

$$S(t) = e^{-\lambda \cdot t}$$

Pero lo más habitual es que la tasa de mortalidad varíe en función del tiempo transcurrido desde la entrada en el estudio.

En ese caso, un modelo muy utilizado, debido a que con la elección adecuada de sus parámetros se ajusta a una gran variedad de situaciones, es el de Weibull:

$$\lambda(t) = c \cdot m \cdot t^{m-1}$$

Modelo de Cox de riesgo proporcional

Es interesante poder modelar no sólo la relación entre la tasa de supervivencia y el tiempo, sino también la posible relación con diferentes variables registradas para cada sujeto. Se trata por tanto de calcular la tasa de mortalidad como una función del tiempo y de las variables pronóstico.

Aunque la idea fundamental es la misma que en cualquier modelo de regresión, aquí la matemática necesaria para la estimación de los coeficientes del modelo se complica sensiblemente, y a pesar de que existen diferentes alternativas, el sistema más utilizado es el denominado de riesgos proporcionales o modelo de Cox, en el que la tasa de mortalidad se calcula como

$$\lambda(t, x_1 .. x_p) = \lambda_0(t) \cdot e^{b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_p \cdot x_p}$$

siendo por tanto el producto de dos componentes, uno $\lambda_0(t)$ que depende del tiempo y otro $e^{b \cdot x}$ que depende de las variables pronóstico o covariantes, y no depende el tiempo.

Vemos que en este modelo los riesgos para dos conjuntos diferentes de valores de los covariantes conservan la misma proporción a lo largo del tiempo; de ahí el nombre de modelos de riesgos proporcionales. Así, por ejemplo, si una de las variables pronóstico fuera tener o no hipertensión, codificada como 0 y 1, manteniendo iguales el resto de covariantes, podemos calcular los valores del término $e^{b \cdot x}$ para sujetos con y sin hipertensión, obteniendo dos números diferentes: 1 para ausencia de hipertensión ($e^0=1$) y si, por ejemplo para los hipertensos fuese 2 , ello supondría según este modelo, que a lo largo del tiempo la tasa de mortalidad de los hipertensos es siempre el doble que para los no hipertensos.

Esta característica del modelo de proporcionalidad de riesgos para diferentes grupos de covariantes hay que tenerla bien presente a la hora de aplicar la técnica a nuestros datos, ya que no siempre es ni siquiera aproximadamente válida tal suposición. Así ocurre cuando la influencia de algún covariante depende precisamente del tiempo.

$\lambda_0(t)$ representa la función de riesgo cuando todos los covariantes valen 0, o el riesgo basal cuando no tiene sentido físico que alguna de las variables valga 0.

Para la estimación de un modelo de riesgos proporcionales se podría postular para la función de riesgo basal un modelo matemático cualquiera, por ejemplo el de Weibull. Sin embargo, la aproximación propuesta por Cox se basa en que a menudo no se conoce la forma de $\lambda_0(t)$ y además no es de interés primordial, ya que el verdadero objetivo es valorar la influencia de los factores pronóstico en la supervivencia. Por ello en el modelo de Cox no se determina $\lambda_0(t)$.

La interpretación de los coeficientes estimados mediante el método de Cox es directa, y se asemeja a la que veíamos en el modelo de regresión logística.

Exp(b_i), donde b_i es el coeficiente correspondiente a la variable X_i, es el **riesgo relativo** cuando X_i aumenta una unidad, manteniéndose constantes las demás.

Precisamente uno de los factores que pueden intervenir en la ecuación es el tratamiento, permitiendo evaluar su influencia en la supervivencia, ante la presencia de otras variables; es decir calcular el efecto del tratamiento en la supervivencia, corrigiendo el efecto debido a otros factores. Cuando únicamente interviene el tratamiento en la ecuación, el resultado es similar (aunque no exactamente igual) a comparar la supervivencia para los diferentes grupos de tratamiento con una prueba como la del logrank descrita en el artículo anterior.

En cuanto a la utilización de **variables cualitativas** con más de dos categorías, valen las consideraciones ya señaladas en el artículo sobre regresión logística.

Presentación de resultados de un modelo de regresión para la supervivencia

Ecuación Regresión

Para cada variable que interviene en la ecuación se indica el coeficiente, su error estándar y el chi² que permite calcular el nivel de probabilidad para ese coeficiente.

Veamos un ejemplo de presentación para un hipotético estudio de supervivencia:

Término	Coef.	Err.est.	Wald chi²	p	Nivel signif.
Colesterol	0.1091	0.0333	10.738	0.0017	$p < 0.01$
Sexo	−0.0488	0.4716	0.011	0.9180	NO
Fumador	1.0638	0.3946	7.268	0.0091	$p < 0.01$

En esta tabla se comprueba que no existe relación entre el sexo y la supervivencia, mientras que los valores de probabilidad sí indican una asociación significativa entre la supervivencia y las otras dos variables: exposición al tabaco y concentración de colesterol. Esto queda mejor ilustrado si se presenta una tabla de riesgos relativos con su intervalo de confianza del 95 %

Riesgo Relativo

Variable	Riesgo Relativo	RR inf.95%	RR sup.95%
Colesterol	1.12	1.04	1.19
Sexo	0.95	0.38	2.40
Fumador	2.90	1.34	6.28

Como se comentó más arriba, el riesgo relativo se calcula como **exp(b_i)**. Un riesgo relativo de 1, corresponde a un coeficiente 0 para esa variable, y si se trata de una característica dicotómica (por ejemplo exposición al tabaco, no o sí) el riesgo es el mismo para los pacientes con o sin la presencia del factor. Un riesgo superior a 1 indica mayor riesgo para los pacientes con esa característica. Así en nuestro caso los pacientes fumadores, a igualdad de otros factores, tienen un riesgo 2.9 veces mayor que los no fumadores.

Una vez más recordemos que el suceso cuya posible aparición en el tiempo se ha estudiado no tiene por qué ser necesariamente la muerte, en este ejemplo podría ser presentar una enfermedad cardiovascular.

Enlaces de interés

- Interesante Calculadora on−line para el cálculo de un modelo de regresión de Cox de riesgos proporcionales, a partir de los datos suministrados por el usuario
- Survival in treated hypertension: follow up study after two decades
  Andersson OV et al. BMJ 1998;317:167−71

Bibliografía seleccionada

- Kleinbaum DG. Statistics in the health sciences: Survival analysis. New York: Springer_Verlag; 1996
- Parmar MKB, Machin D. Survival analysis: a practical approach. Chichester: Wiley.

Indice de artículos   Principio de la página