Más allá de los modelos de Cox: modelos paramétricos de supervivencia

Preparado por Luis M. Molinero (Alce Ingeniería)
Octubre 2004
CorreoE: **bioestadistica@alceingenieria.net**

Artículo en formato PDF
www.seh−lelha.org/stat1.htm

Introducción

Los modelos de riesgo proporcional conocidos como modelos de Cox son sin lugar a dudas los más utilizados en la investigación clínica para analizar datos de supervivencia, y por tanto los que más frecuentemente aparecen en la literatura médica. Recordemos que en este tipo de modelos la función de riesgo tiene la siguiente forma:

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \qquad [1]$$

Una característica de los modelos de Cox es que no se especifica la forma de la función de riesgo base $h_0(t)$. Esto puede tener sus ventajas, o por el contrario ser un inconveniente. Cuando el objetivo es comparar grupos, valorar supervivencias relativas, lo que interesa es calcular cocientes de riesgo y al dividir las dos funciones, como el término $h_0(t)$ interviene en ambas, desaparece, por lo que en estos casos realmente da igual cual pueda ser la forma de esta función.

Sin embargo esta característica puede ser un inconveniente cuando lo que se desea es calcular un valor absoluto de supervivencia para un determinado perfil de riesgo, ya que entonces sí es necesario estimar $h_0(t)$, y si dicha estimación se efectúa directamente a partir de los datos, como en los modelos de Cox, se obtiene una función que cambia para cada tiempo en el que se observa algún suceso, por lo que en realidad es muy dependiente de los datos concretos que hemos obtenido en nuestro estudio. Es como si en una regresión en lugar de ajustar a los datos la ecuación de una recta o de una parábola, estimáramos un polinomio de mayor grado que pasase por todos los puntos. Por otro lado, si se desea utilizar el modelo con finalidad predictiva, con los modelos de Cox no se puede proporcionar una fórmula compacta para el cálculo de la supervivencia ya que $S_0(t)$ no queda expresado por una fórmula, sino por una función con forma escalonada, cuyo valor cambia para cada tiempo en el que se observa un suceso.

Es en estos casos cuando la utilización de modelos de probabilidad específicos (**modelos paramétricos**) para la función de riesgo puede ser de gran interés. Además si la utilización de una distribución de probabilidad concreta es adecuada, las inferencias basadas en estos modelos son más precisas.

En este artículo vamos a comentar un tipo de distribución muy utilizada en supervivencia, que se conoce con el nombre de **distribución de Weibull**. Precisamente este tipo de modelo paramétrico, basado en la distribución de Weibull, es el que se ha empleado en el reciente proyecto SCORE.

Vamos también a proporcionar un ejemplo de datos de supervivencia con cáncer de pulmón (Prentice 1973), con el fin de que el lector pueda verificar con su software estadístico favorito los cálculos aquí presentados.

Distribución de Weibull

El modelo de función de riesgo denominado de Weibull viene dado por la siguiente expresión

$$h(t) = \lambda\gamma\ t^{\gamma-1} \qquad [2]$$

que depende de dos parámetros λ y γ, ambos mayores que cero. La forma de la función de riesgo viene dada por γ. En el caso concreto de γ=1 la función de riesgo es constante, independiente del tiempo, con valor λ, lo que corresponde a una función de supervivencia **exponencial**. Para γ<1 la función de riesgo disminuye con el tiempo (puede servir por ejemplo para modelar la aparición de fallos en el software, que no envejece, salvo Windows claro), y para γ>1 la función de riesgo aumenta con el tiempo (por ejemplo para modelar la supervivencia del hardware, que envejece con el tiempo).

La supervivencia correspondiente a la función de riesgo de Weibull viene dada por la siguiente fórmula:

$$S(t) = \exp(-\lambda t^{\gamma}) \qquad [3]$$

que como vemos no es demasiado complicada.

Para λ =1 en la figura 1 se representa la función de riesgo para tres valores de γ, y en la figura 2 las curvas de supervivencia respectivas:

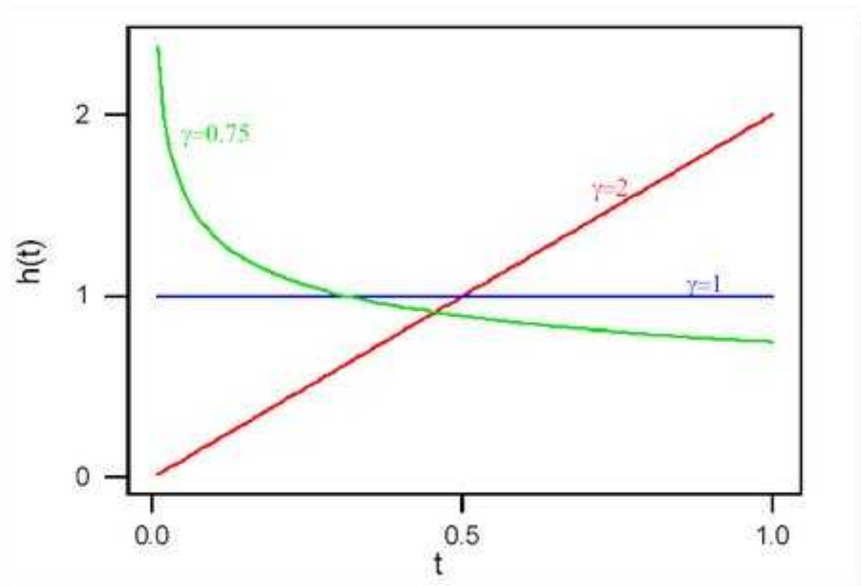

Fig.1 Función de riesgo Weibull

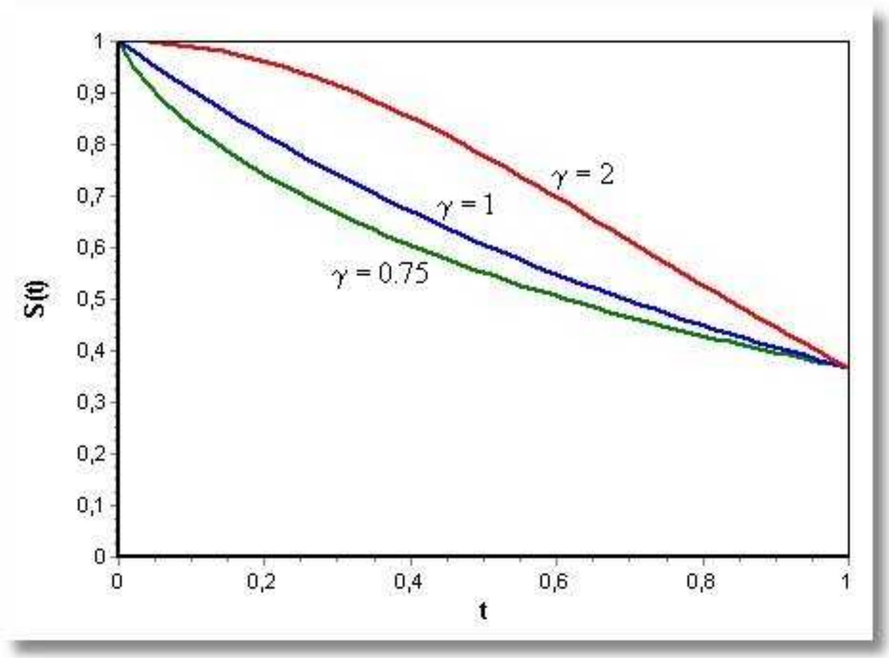

Fig.2 Función de Supervivencia Weibull

Verificación del ajuste de un modelo de Weibull

Para verificar si un modelo de Weibull es adecuado a nuestros datos disponemos de un método gráfico bastante simple. Si en la fórmula [3] realizamos unas sencillas manipulaciones algebraicas obtenemos esta otra fórmula:

$$\ln[-\ln S(t)] = \ln \lambda + \gamma \ln t \qquad [4]$$

que es la ecuación de una recta función de ln t. Luego si representamos en una gráfica ln[−ln S(t)] en función de ln t y el modelo de Weibull es adecuado, los puntos deberán estar próximos a una línea recta.

Si llevamos a cabo esa representación para los datos del ejemplo obtenemos la siguiente gráfica:

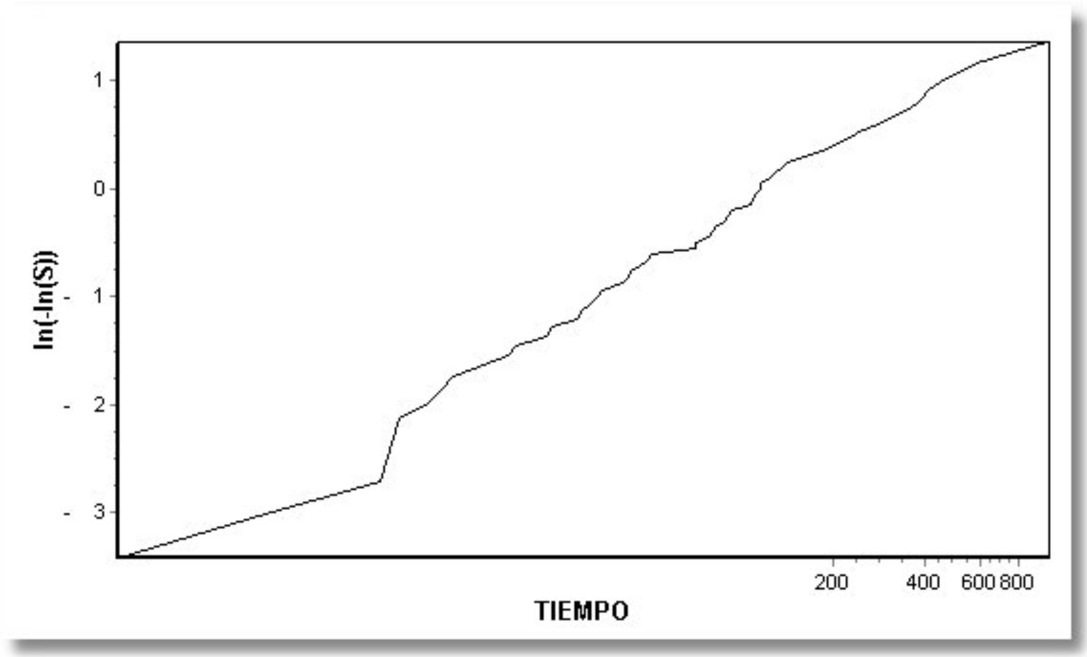

Fig.3

que no se aleja mucho de una línea recta (salvo quizás en la zona inicial), por lo que en este caso un modelo de Weibull puede ser adecuado, como también podemos comprobar si se representa en el mismo gráfico la curva de supervivencia estimada por el modelo de Kaplan−Meier y el modelo de Weibull:

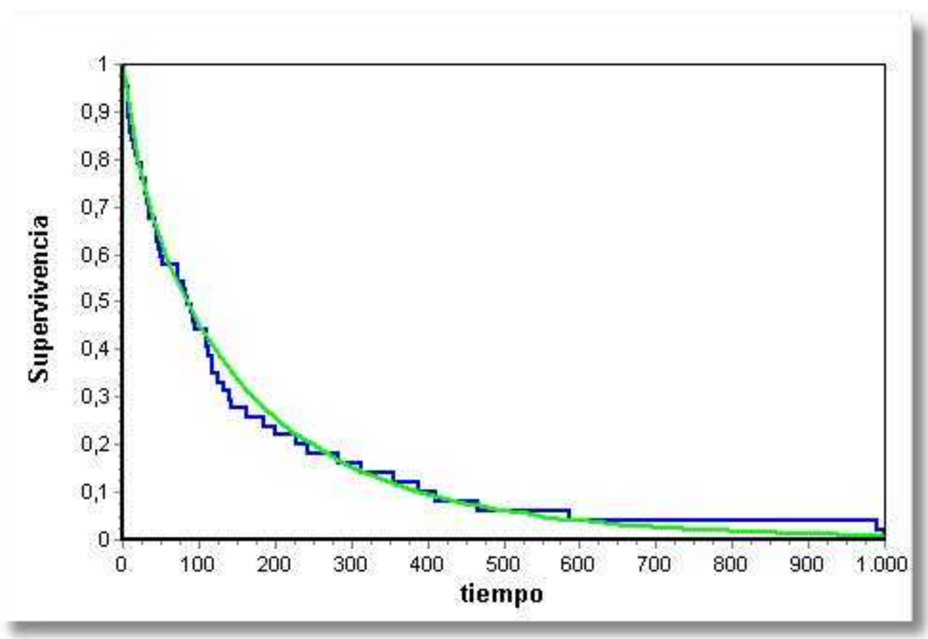

Fig.4 La línea verde corresponde al modelo de Weibull y la línea azul a la supervivencia estimada por el método de Kaplan−Meier

Si estimamos un modelo de Weibull de supervivencia para los [datos del ejemplo] se obtiene un valor de γ=0.781 y un valor de λ=0.0217

Hay que tener en cuenta que en ocasiones el modelo de Weibull se expresa de forma algo diferente a la presentada en la fórmula [2] mediante la siguiente expresión

$$h(t) = \alpha\gamma(\alpha t)^{\gamma-1} \qquad [5]$$

En ambas fórmulas el valor de γ es el mismo y $\lambda = \alpha^{\gamma}$. Luego si el programa utilizado hubiera estimado el modelo según esta fórmula habríamos obtenido el mismo valor de γ y un valor de α=0.00742

Modelo de riesgos proporcionales de Weibull

En el modelo de Weibull podemos también introducir variables predictoras.

Si se supone que la función de riesgo para dos perfiles diferentes se mantiene proporcional a lo largo del tiempo, la función de riesgo para el sujeto i en función de p variables predictivas es

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \qquad [6]$$

Aunque se trata de la misma fórmula que [1], correspondiente al modelo de riesgos proporcionales de Cox, la diferencia radica en que ahora para la función de riesgo base $h_0(t)$ suponemos que sigue un modelo de Weibull, mientras que en la fórmula de Cox no se hacía ninguna suposición respecto a la forma particular de esta función. Entonces la función de riesgo y de supervivencia vienen dadas por las siguientes expresiones:

$$h_i(t) = \exp(\beta x_i)\lambda\gamma t^{\gamma-1} \qquad [7]$$

$$S_i(t) = \exp[-\exp(\beta x_i)\lambda t^{\gamma}] \qquad [8]$$

La interpretación de los coeficientes β es idéntica a la que se efectúa en el modelo de riesgo proporcional de Cox, y por tanto también ahora los términos exp(β) son análogos al **riesgo relativo**.

Hay que resaltar que la mayoría de los programas de ordenador ajustan el modelo de riesgo proporcional de Cox con otra expresión, que no vamos a entrar a comentar aquí para no complicar la exposición. Habitualmente los programas nos proporcionan además de los coeficientes (que por ser diferentes de los del modelo aquí presentado vamos a llamar α), dos parámetros que suelen denominar *intercept* y *scale*. La relación entre los valores proporcionados por los programas y el modelo aquí presentado es la siguiente:

$$\lambda = \exp(-intercept\,/\,scale)$$

$$\gamma = 1/\,scale$$

$$\beta = -\,\alpha\,/\,scale$$

A continuación se indica una salida típica de un programa cuando se estima un modelo de Weibull para los datos del ejemplo, considerando como factor pronóstico (covariante) el tratamiento (hay dos grupos de tratamiento):

```
---------------------------------------------------------
             Value Std. Error     z        p
(Intercept) 4.647      0.268 17.32 3.52e-67
TRATAMIEN   0.410      0.342  1.20 2.31e-01
Log(scale)  0.225      0.102  2.20 2.80e-02

Scale= 1.252

Weibull distribution
Loglik(model)= -339.7   Loglik(intercept only)= -340.4
        Chisq= 1.35 on 1 degrees of freedom, p= 0.24
Number of Newton-Raphson Iterations: 5
n= 62
---------------------------------------------------------
```

Por lo tanto el valor de λ =exp(−4.647 / 1.252)=0.0243

γ=1 / 1.252 = 0.799

β=−0.41 / 1.252 = −0.327

Exponenciando β obtenemos el riesgo relativo = 0.721

Mientras que si ajustamos un modelo proporcional de Cox a esos mismos datos, veremos que obtenemos un valor de β=−0.247 que corresponde a un riesgo relativo de 0.78.

Modelo log−logístico

Una característica presente en los modelos de Weibull es que la función de riesgo, dependiendo de los valores de los parámetros, es o bien creciente o bien decreciente (salvo el caso particular de λ =1 para el que es constante). Sin embargo en la vida real a veces la función de riesgo puede variar su sentido. Por ejemplo, después de una intervención quirúrgica es posible que inicialmente el riesgo sea alto y posteriormente, pasada una fase crítica, el riesgo disminuya. En estos casos no nos sirven los modelos de Weibull, pero existen otras alternativas. La más empleada es el modelo **log−logístico**, cuya función de riesgo y de supervivencia tienen las siguientes expresiones

$$h(t) = \frac{e^{\theta}\kappa t^{\kappa-1}}{1+e^{\theta}t^{\kappa}} \quad [9]$$

$$S(t) = \frac{1}{1+e^{\theta}t^{\kappa}} \quad [10]$$

En la figura se representa una función de riesgo construida con este modelo, que crece inicialmente y a partir de un momento determinado cambia de sentido y decrece.

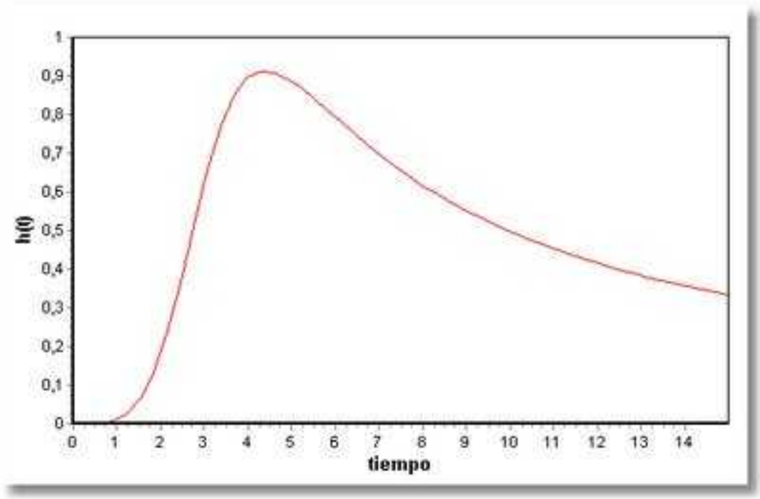

Fig.5 Función de riesgo para un modelo log−logístico con κ =5 y θ=−6

También aquí, con unas sencillas manipulaciones algebraicas, podemos obtener la expresión de la ecuación de una recta, que nos permite verificar gráficamente la adecuación de nuestros datos al modelo:

$$\ln\left[\frac{S(t)}{1-S(t)}\right] = -\theta - \kappa \ln t \qquad [11]$$

Por lo tanto si se representa el log−odds ln [S/(1−S)] de la supervivencia en función de ln t y se puede considerar que los puntos no se alejan de una recta, será adecuado utilizar este modelo.

Todo lo comentado para los modelos proporcionales de Weibull se puede aplicar a este modelo.

Epílogo

Cuando se desea estimar un modelo predictivo de supervivencia cuyo objetivo va a ser calcular valores absolutos de supervivencia resulta interesante disponer de alternativas paramétricas a los ubicuos modelos de Cox, ya que nos proporcionan la posibilidad de emplear una fórmula más compacta y por lo tanto fácilmente programable. Aunque no se utilizan excesivamente en la literatura médica, resulta ilustrativo el que precisamente en uno de los modelos recientes de predicción de riesgo cardiovascular −proyecto SCORE− haya sido empleado un modelo de Weibull.

Referencias

- **Modelling Survival Data in Medical Research**
  David Collet
  Chapman&Hall/CRC. 2003
