

Selección de modelos de supervivencia en la industria farmacéutica¹

Rafael E. Borges²

Departamento de Estadística, Universidad de Los Andes, Mérida, Venezuela

1. Introducción

El propósito de este trabajo es alertar sobre posibles problemas que pudieran presentarse al utilizar el modelo de Cox y proponer modelos alternativos en casos de observarse violación de los supuestos u otros problemas más difíciles de identificar. Se presenta además un ejemplo aplicado a la industria farmacéutica.

2. Elementos del análisis de supervivencia

El análisis de supervivencia engloba un conjunto de técnicas para analizar el tiempo de seguimiento hasta la ocurrencia de un evento de interés, tiempo que puede observarse completa o parcialmente. La observación parcial puede ser debida a mecanismos de censura y trucamiento.

Sea T una variable aleatoria positiva (o no negativa) con función de distribución $F(t)$ y función de densidad de probabilidad $f(t)$. La función de supervivencia $S(t)$ se define como:

$$S(t) = 1 - F(t) = P[T > t]$$

Otra función importante es la función de razón de riesgos o tasa instantánea de fallas $\lambda(t)$, definida como la la probabilidad de que a un individuo le ocurra el evento de interés en la

¹Trabajo realizado con apoyos del Consejo de Desarrollo Científico, Humanístico y Tecnológico de la Universidad de Los Andes (CDCHT-ULA), Mérida, Venezuela, a través del proyecto E-199-02-09-C.

²borgesr@ula.ve

siguiente unidad de tiempo Δt dado que ha sobrevivido hasta el tiempo t .

En el análisis de supervivencia, el primer paso, consiste en estimar la función de supervivencia, para lo cual existen varios estimadores, siendo el más utilizado, el estimador de Kaplan y Meier (1958).

Otro aspecto importante es la utilización de modelos de regresión, que sean capaces de explicar la función de riesgo pudiendo incluirse variables explicativas, denominadas, covariables. Algunos de estos modelos serán presentados en las próximas secciones.

2.1. El modelo de regresión de Cox

El modelo de regresión de Cox (1972) es el modelo de regresión más utilizado para datos de supervivencia en el área médica. En el modelo de regresión de Cox, el riesgo para el i -ésimo individuo se define mediante:

$$\lambda(t; Z_i(t)) = \lambda_0(t) e^{\beta' Z_i(t)}$$

donde $Z_i(t)$ es el vector de covariables para el i -ésimo individuo en el tiempo t .

El ajuste del modelo de Cox se hace a través del método de verisimilitud parcial. Una vez que se ha ajustado un modelo de Cox, existen tres contrastes de hipótesis para verificar la significación del modelo, estos tests son asintóticamente equivalentes pero no siempre sucede lo mismo en la práctica. Estos contrastes son el de razón de verosimilitud, el de Wald y el de los puntajes (score test).

2.2. Análisis de residuos, violación de supuestos, modelos alternativos y criterios para la selección de modelos

Una de las ventajas que han surgido del enfoque del análisis de supervivencia es la posibilidad de efectuar análisis de residuos (Therneau y Grambsch, 2000). Los residuos se pueden utilizar para verificar el supuesto de riesgo proporcional (supuesto más importante) (residuos de

Schoenfeld), para identificar los individuos influyentes en la estimación del modelo (residuos deviance), para identificar los individuos influyentes en la estimación del parámetro asociado a cada covariable (residuos score) y para descubrir la forma funcional correcta de un predictor continuo (residuos de martingala).

Existen otras situaciones en las que, a pesar de la no violación de los supuestos, pareciera ser necesario utilizar otros modelos. Por ejemplo, para los casos en que se crucen las funciones de supervivencia para dos más grupos se pudiera utilizar un modelo de Cox Generalizado (Hsieh, 2001, Bagdonavicius y Nikulin, 2001), un patrón quebradizo en el gráfico para verificar el supuesto de riesgo proporcional pudiera sugerir el uso de modelos locales de Cox y, algunos patrones cíclicos en este gráfico sugieren la posibilidad de utilizar modelos de Cox con covariables en el tiempo (Vaupel *et al*, 1979).

Otros modelos alternativos son el modelos aditivo de Aalen (1980), el modelo de Cox-Aalen y otros modelos dinámicos (Martinussen y Scheike, 2006), el modelo de Gray (1992), el modelo de regresión de Buckley y James (1979), los modelos acelerados (Bagdonavicius y Nikulin, 2001).

Exite otra clase de modelos basados en familias paramétricas (exponencial, Weibull, logística, normal, log-logística, log-normal, valor extremo, entre otras), algunas de las cuales pueden ser sugeridas mediante gráficos de probabilidad o mediante algunos gráficos específicos (Meeker y Escobar, 1998).

Un aspecto importante a considerar es que cada modelo tiene unos supuestos que deben ser verificados. Pudiera además presentarse el caso de tener varios modelos, y para la selección del mejor de ellos pueden utilizarse varios criterios, siendo el más utilizado el criterio de información de Akaike.

3. Ejemplo

En esta sección se presenta un ejemplo aplicado a la industria farmacéutica. Los datos han sido tomados de un informe de asesoría llevada a cabo por Borges (2004).

El análisis efectuado consistió en la estimación de la función de supervivencia mediante el estimador de Kaplan y Meier, y el ajuste del modelo de Cox, incluyendo además la verificación de los residuos.

El modelo de Cox obtenido incluía a las covariables DOSIS y POSTDATISMO. Tanto el modelo como las variables resultaron significativas y no se encontró violación de los supuestos del modelo.

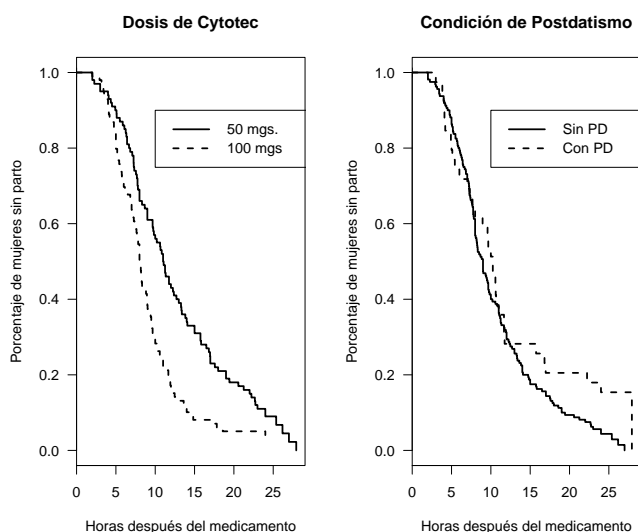


Gráfico 1. Función de Supervivencia estimada a través del estimador de Kaplan y Meier.

En el gráfico 1 se observa, por una parte que el medicamento reduce significativamente el tiempo de trabajo de parto (resultado similar al del modelo de Cox) y, el entrecruzamiento de las curvas para el postdatismo, indica que el modelo de Cox no es del todo adecuado, por lo que debería ajustarse otro modelo (modelo de Cox generalizado o modelo aditivo de Aalen).

4. Referencias

Aalen, O.O. (1980). A Model for Non-parametric Regression Analysis of Counting Processes. In *Lecture Notes in Statistics 2* (eds. W. Klonecki *et al*), pp. 1-25. New York: Springer-Verlag.

- Bagdonavicius, V. y Nikulin, V. (2001). *Accelerated Life Models: Modeling and Statistical Analysis*. London: Chapman & Hall.
- Borges, R. (2004). Estudio Comparativo de dos dosis del Medicamento Cytotec como Potencial Fármaco para la Inducción del Parto, *Informe Técnico de Asesoría*, Mérida, Venezuela: Grupo de Investigación en Bioestadística de la Universidad de Los Andes.
- Buckley, J.J. y James, I.R. (1979). Linear regression with censored data, *Biometrika*, **66**, 429-436.
- Cox, D.R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society: Series B*, **34**, 187-220.
- Gray, R.J. (1992). Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis, *Journal of the American Statistical Association*, **87**, 942-951.
- Hsieh, F. (2001). On heteroscedastic hazard regression models: theory and application, *Journal of the Royal Statistical Society: Series B*, **61**, 63-79.
- Kaplan, E.L. y Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457-481.
- Martinussen, T. y Scheike, T.H. (2006). *Dynamic Regression Models for Survival Data*. New York: Springer-Verlag.
- Meeker, W.Q. y Escobar, L.A (1998). *Statistical Methods for Reliability Data*. New York: Wiley.
- Therneau, T.M. y Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.
- Vaupel, J., Manton, K. y Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, **16**, 439-454.