



Revista Española de Salud Pública

ISSN: 1135-5727

resp@msc.es

Ministerio de Sanidad, Servicios Sociales e
Igualdad
España

González, Juan R.; Peña, Edsel A.

Estimación no paramétrica de la función de supervivencia para datos con eventos recurrentes

Revista Española de Salud Pública, vol. 78, núm. 2, marzo-abril, 2004, pp. 189-199

Ministerio de Sanidad, Servicios Sociales e Igualdad

Madrid, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=17078206>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

COLABORACIÓN ESPECIAL**ESTIMACIÓN NO PARAMÉTRICA DE LA FUNCIÓN DE SUPERVIVENCIA
PARA DATOS CON EVENTOS RECURRENTE (*)****Juan R. González (1) y Edsel A. Peña (2)**

(1) Servicio de Prevención y Control del Cáncer. Institut Català d'Oncologia. Barcelona.

(2) Department of Statistics. University of South Carolina. Columbia. USA

RESUMEN

Los eventos recurrentes cuando tratamos estudios de supervivencia, necesita utilizar una metodología distinta a la empleada en el análisis de supervivencia estándar. El principal problema que nos encontramos a la hora de realizar inferencia en este tipo de estudios es que las observaciones pueden no ser independientes. Así, si no tenemos en cuenta este hecho, se pueden obtener estimadores sesgados e ineficientes. En el caso de independencia, podemos usar la generalización del estimador límite del producto propuesto por Peña et al. (2001) para estimar la función de supervivencia de los tiempos de interocurrencias. Sin embargo para el modelo con tiempos correlacionados debemos utilizar otros modelos como el modelo de fragilidad (frailty model) o un estimador propuesto por Wang y Chang (1999) que contempla tanto el hecho que los tiempos estén o no correlacionados. El objetivo de este trabajo ha sido ilustrar estas aproximaciones con dos ejemplos basados en datos reales.

Palabras clave: Análisis de supervivencia. Recurrencias. Estadística no paramétrica. Readmisiones.

ABSTRACT**Nonparametric Estimation of Survival
Function with Recurrent Event Data**

Recurrent events when we deal with survival studies demand a different methodology from what is used in standard survival analysis. The main problem that we found when we make inference in these kind of studies is that the observations may not be independent. Thus, biased and inefficient estimators can be obtained if we do not take into account this fact. In the independent case, the interoccurrence survival function can be estimated by the generalization of the limit product estimator (Peña et al. (2001)). However, if data are correlated, other models should be used such as frailty models or an estimator proposed by Wang and Chang (1999), that take into account the fact that interoccurrence times were or not correlated. The aim of this paper has been the illustration of these approaches by using two real data sets.

Key words: Survival analysis. Recurrences. Nonparametric statistics. Patient readmission.

INTRODUCCIÓN

Los estudios de supervivencia se interesan por el tiempo que transcurre desde el inicio del seguimiento (nacimiento, diagnóstico de una enfermedad, inicio del tratamiento,...) hasta que ocurre un suceso de interés

(muerte, curación, mejora, ...)¹. Sin embargo, en muchas ocasiones este suceso se observa más de una vez en un mismo individuo a lo largo del periodo de seguimiento. Tanto en el ámbito biomédico, de salud pública o de ensayos clínicos, podemos encontrar numerosos ejemplos en los que el

Correspondencia:
Juan R. González
Servicio de Epidemiología y Registro del Cáncer
Institut Català d'Oncologia
Avda. Gran Vía s/n, km 2,7
L'Hospitalet de Llobregat
08907-Barcelona, Spain
Correo electrónico: jrgonzalez@ico.scs.es

(*) Este trabajo ha sido financiado en parte por las Redes de Temáticas de Investigación Cooperativa en Cáncer (C03/10) y en Epidemiología y Salud Pública (C03/09), financiadas por el Instituto Carlos III, Ministerio de Sanidad y Consumo. El trabajo del segundo autor está parcialmente financiado por NSF Grant DMS 0102870.

evento analizado aparece de manera recurrente: la recaída de un tumor después de su cirugía, los ataques epilépticos, el abuso de drogas o alcohol en adolescentes, los dolores de cabeza recurrentes (migrañas), las hospitalizaciones por enfermedades renales o por cáncer, el movimiento del intestino delgado durante la ingesta, el inicio de una depresión, las fracturas en pacientes con cáncer metastásico en los huesos, la angina de pecho en pacientes con enfermedades coronarias, o las enfermedades cardiovasculares. Es claro que si estamos interesados en estudiar las propiedades de la función de distribución del tiempo de interocurrencia de estos eventos, debemos tener en cuenta la naturaleza recurrente de los mismos. Existen muchos aspectos que debemos tener presentes a la hora realizar dicha estimación, pues ésta vendrá influenciada evidentemente por el tiempo hasta que observamos el evento, pero también por el número de ocurrencias y, de la misma forma que en el caso de un evento único, por la censura producida generalmente por el fin del estudio.

En este trabajo se expondrán los diferentes aspectos que se deben estudiar a la hora de estimar la función de supervivencia en presencia de eventos recurrentes, así como algunas de las aproximaciones estadísticas de las que se disponen para realizar dicha estimación. Finalmente, esta metodología se ilustrará con dos ejemplos con datos reales en los que se discutirán estos problemas. Además, se ilustrará cómo utilizar el software estadístico existente.

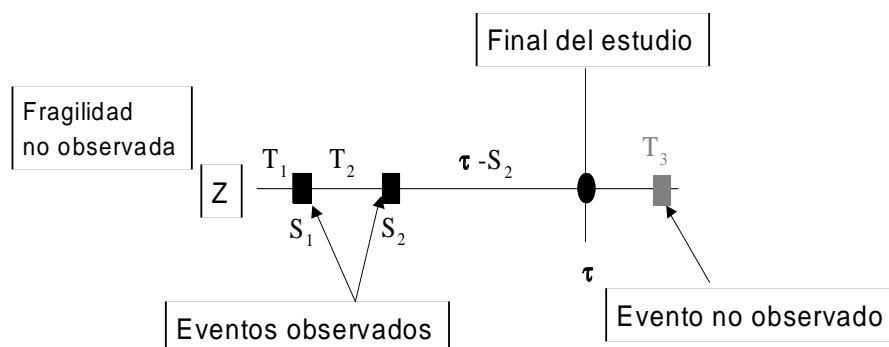
MÉTODOS

Notación

La figura 1 representa una situación habitual en los estudios en los que se observa un evento de forma recurrente. Para cada uno de los individuos disponemos de un periodo de seguimiento $[0, \tau]$ donde τ puede representar o bien el tiempo final del estudio o bien el tiempo hasta que podemos seguir al individuo. Este tiempo τ puede estar gobernado por una función de probabilidad desco-

Figura 1

Esquema de un estudio de supervivencia para eventos recurrentes. S hace referencia al tiempo calendario y T al tiempo de interocurrencias o de brecha (*gap time*)



nocida $G(\tau) = P(\tau \leq t)$. Los eventos ocurren en los tiempos calendarios $S_0 \equiv 0 < S_1 < S_2 < \dots$ y los tiempos de interocurrencia o de brecha (gap time en inglés) T_1, T_2, T_3, \dots nos identifica el tiempo entre dos eventos consecutivos (ver figura 1). Así, para $i=1, 2, 3, \dots$ tenemos que

$$T_i = S_i - S_{i-1} \text{ y } S_i = T_1 + T_2 + \dots + T_i$$

Además durante el periodo de observación $[0, \tau]$, el número de interocurrencias viene dado por

$$K = \max \{k \in \{0, 1, 2, \dots\} : S_k \leq \tau\},$$

que también es una variable aleatoria y cuya distribución depende tanto de la distribución de los tiempos de interocurrencias T_i como de la distribución G de τ . En resumen, para el individuo i observamos

$$D_i = (K_i, T_{i1}, T_{i2}, \dots, T_{iK_i}, \tau_i - S_{K_i})$$

La especificación del último tiempo $\tau - S_K$ es redundante ya que queda determinado tras observar los K tiempos de interocurrencias. Sin embargo, incluiremos la notación $\tau - S_K$ ya que este valor corresponde al tiempo censurado por la derecha del tiempo de interocurrencia T_{K+1} . Por ejemplo, en la figura 1 el tiempo $\tau - S_2$ corresponde al tiempo censurado para el evento 3. En este caso T_3 no se observa completamente debido a que el tercer evento se produciría con posterioridad a la finalización del estudio. Cuando consideremos esta estructura de los datos en la situación de eventos recurrentes, es necesario tener en cuenta que tanto K como el mecanismo de censura para T_{K+1} son informativos²⁻⁴.

Estimación de la función de supervivencia

La estimación de la función de supervivencia (o de distribución) para el caso de eventos recurrentes ha sido considerada por muchos autores^{2,4-9}. La caracterización de la función de distribución de los tiempos de

interocurrencias viene claramente determinada por el esquema de tiempos que se ha mostrado con anterioridad. Peña et al. asumen que los tiempos de interocurrencias representan observaciones independientes e idénticamente distribuidas (i.i.d.) de una función de distribución desconocida F , y que cada individuo es observado en un periodo de tiempo aleatorio⁴. Este modelo es razonable en ingeniería y fiabilidad, pero en el ámbito biomédico es a menudo demasiado restrictivo ya que: (i) la distribución del tiempo al primer evento puede diferir de la distribución de los tiempos de interocurrencias; (ii) los tiempos de interocurrencias pueden estar correlacionados; y (iii) los tiempos de interocurrencias pueden depender de covariables. Para el problema (i) Wang y Chang² y Peña et al.⁴ resuelven el problema asumiendo que el evento inicial es el criterio de admisión en el estudio. Para el problema (ii) se pueden utilizar los estimadores, que se presentarán más adelante, propuestos por Wang y Chang o por Peña et al. bajo el caso que los tiempos de interocurrencias sigan un modelo gamma de fragilidad (gamma frailty model en inglés). Finalmente, diversas extensiones basadas en el modelo de riesgos proporcionales de Cox¹⁰⁻¹² (modelos marginales y condicionales) se pueden utilizar para resolver el problema (iii). Puede leerse una buena revisión de estos modelos publicada en castellano recientemente¹³.

Modelo de fragilidad

Una asunción que se hace muy a menudo cuando se modela el efecto de un tratamiento o un potencial factor de riesgo en la supervivencia, es que los tiempos de eventos de los individuos son, condicionalmente a las variables observadas, estadísticamente independientes. En la práctica, puede darse el caso que los tiempos de eventos, de los individuos en algunos subgrupos estén asociados ya que los miembros de esos grupos tienen un trato común no observado. Este es quizás el mayor problema metodológico en el contexto de los eventos recurrentes (pro-

blema (ii)), es decir, cómo tratar la dependencia de los tiempos de interocurrencia. En estudios con humanos puede existir una asociación en los tiempos de eventos como cáncer, enfermedades cardiovasculares, o en la muerte, entre gemelos y/o entre familiares. En un contexto más general, esta asociación podría darse en pacientes con una misma predisposición hereditaria o genética. Un modelo específico que se suele utilizar cuando se tiene correlación o dependencia en los tiempos de interocurrencias, es el modelo multiplicativo de fragilidad^{14,15} (*multiplicative frailty model* en inglés). El término fragilidad se entiende como heterogeneidad individual, en el sentido que hay un grupo de individuos más «frágiles» que otros que hacen que el evento se observe en ellos con una mayor probabilidad. A veces esta variable puede controlarse (caso de las familias) pero en general, son difíciles de medir (información sobre susceptibilidad genética o predisposición hereditaria). También puede darse el caso que esta heterogeneidad sea debida a variables medibles pero que no han sido recogidas por el investigador durante el estudio. Bajo este modelo, se postula que para cada sujeto existe una variable de fragilidad no observable positiva Z_i tal que, condicionalmente a $Z_i = z$, los tiempos de interocurrencias T_1, T_2, \dots son i.i.d. con una función condicional de supervivencia, S , común

$$S(t | Z_i = z) = [S_0(t)] = \exp\left(-z \int_0^t \lambda_0(u) du\right)$$

donde $\lambda_0(\cdot)$ es la función de riesgo asociada a la función de supervivencia basal $S_0(\cdot)$. Las fragilidades Z_1, Z_2, \dots, Z_n se asumen que son i.i.d. de una distribución desconocida H . En general, las Z no se observan por lo que estamos interesados en estimar la función de supervivencia marginal de T_{ij} que bajo este modelo se expresa como

$$S(t) = \text{Esperanza}\{\exp(-Z_1 \Lambda_0(t))\}$$

donde $\Lambda_0(t) = -\log [S_0(t)]$ es la función de riesgo acumulada de S_0 .

Una elección muy común de la función de distribución de fragilidad H , es una distribución gamma con parámetros de forma y escala igual a α . En este caso, la distribución de supervivencia bajo el modelo de fragilidad puede escribirse como

$$S(t) = \left[\frac{\alpha}{\alpha + \Lambda_0(t)} \right]^\alpha. \quad (1)$$

El parámetro α controla el grado de asociación de los tiempos de ocurrencias dentro de cada sujeto. En particular, si α crece (decrece), la asociación entre los tiempos de interocurrencias decrece (crece).

Estimador Peña, Strawderman y Hollander

Peña et al.⁴ desarrollan el estimador no paramétrico de máxima verosimilitud bajo la asunción del modelo i.i.d. Generalizan el estimador límite del producto (GPLe) o de Kaplan-Meier para el caso con eventos recurrentes mediante el uso de procesos contadores. En general, esta formulación define dos funciones, N e Y , que en el caso particular de supervivencia con una única ocurrencia por individuo, indican el número de eventos que ocurren en un periodo de tiempo y los individuos a riesgo para un tiempo dado, respectivamente (puede verse la introducción del libro de Therneau y Grambsch¹⁶ para una explicación muy intuitiva de esta metodología). En el caso de eventos recurrentes los procesos que se definen están doblemente indexados ya que se trabaja con dos escalas de tiempo: el tiempo calendario s y el tiempo de interocurrencia t ^{4,17,5}. En este caso los procesos N e Y se definen simultáneamente en la escala de tiempo calendario y en el de interocurrencia. Así, $N(s, t)$ cuenta el número de eventos observados que ocurren para un periodo de tiempo $[0, s]$ cuyos tiempos de tiempos de interocurrencia son como mucho t , e $Y(s, t)$ cuenta el número de eventos obser-

vados en el tiempo calendario $[0, s]$ cuyos tiempos de interocurrencia son como mínimo t . En el Anexo A puede verse una definición más formal de estos procesos.

Así, mediante la definición de estos nuevos procesos, el estimador GPLE se expresa como

$$\hat{S}(s, t) = \prod_{w \leq t} \left[1 - \frac{N(s, \Delta w)}{Y(s, w)} \right]$$

Peña et al también proponen un estimador bajo el modelo gamma de fragilidad (referido como FRMLE en el artículo), y muestran como la estimación de α y Λ_0 en (1) puede obtenerse mediante la maximización de la función de verosimilitud marginal de α y Λ_0 (.) y utilizando el algoritmo EM⁴. El estimador de (1) es de la forma

$$\tilde{S}(s, t) = \left[\frac{\hat{\alpha}}{\hat{\alpha} + \hat{\Lambda}_0(s, t)} \right]^{\alpha}$$

donde $\hat{\Lambda}_0(s, t)$ es un estimador de la función de riesgo acumulada marginal de $\Lambda_0(t)$.

Estimador Wang y Chang

Wang y Chang² (WC) proponen otro estimador para la función de supervivencia marginal en el caso que exista correlación entre los tiempos de ocurrencias. Ellos consideran una estructura de correlación más general, que incluye como un caso especial el modelo i.i.d. y el modelo de fragilidad gamma. Su estimador puede definirse utilizando dos nuevos procesos, d^* y R^* . Los autores intentan tener en cuenta en la definición de N e Y que un individuo pueda tener más de un evento. De hecho, este estimador tiene la misma forma que el estimador GPLE pero utilizando estos dos nuevos procesos. Así, d^* representa la suma de la proporción de individuos de los tiempos de interocurrencia que son iguales a t cuando hay al menos un

evento. Por otro lado, R^* representa una media de los individuos que están a riesgo a tiempo t , donde para cada individuo la media es el número de fallos o tiempos censurados al menos iguales a t . Este promedio se hace respecto al número de eventos que hay para cada individuo y en caso que K sea 0 se divide por 1 (ver anexo 1 para una definición más formal). Utilizando esta notación, el estimador WC de S , si denotamos a \mathfrak{T} al conjunto de tiempos distintos de interocurrencias para los n individuos, viene dado por

$$\tilde{S}(t) = \prod_{(T_k \in \mathfrak{T}; T_k \leq t)} \left[1 - \frac{d^*(T_k)}{R^*(T_k)} \right].$$

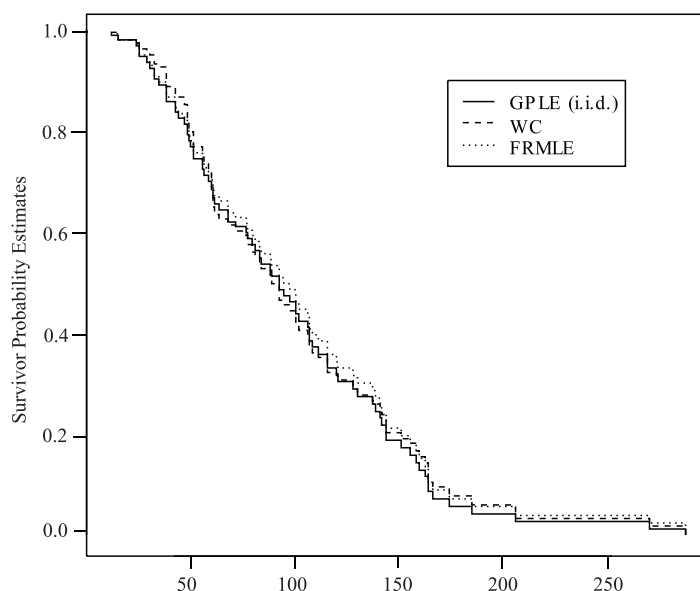
Estos tres estimadores pueden calcularse utilizando la librería para R llamada *survrec*¹⁸ que está disponible gratuitamente en <http://www.cran.r-project.org>. A continuación se ilustrará el cálculo de la función de supervivencia mediante estos estimadores en dos ejemplos con datos reales. Las funciones para obtener los resultados y las gráficas pueden verse en el anexo 2.

Ejemplos ilustrativos

Ejemplo 1: El primer ejemplo hace referencia a un estudio sobre motilidad del intestino delgado¹⁹. El principal objetivo de los investigadores fue estimar el tiempo medio del periodo del Complejo Migratorio Motor (CMM). Los autores consideran dos aproximaciones: la primera basada en un modelo de componentes de la varianza, y la segunda utilizando un modelo de Weibull con una componente gamma de fragilidad. Los autores asumen el modelo de independencia (es decir, que los tiempos de interocurrencias ocurren de forma independientemente, sin correlación). Para ello argumentan que: «los periodos de CMM individuales aparecen aproximadamente como un proceso de renovación». Sin embargo, no utilizan ningún método estadístico para testar dicha hipótesis. Como veremos en el ejemplo 2, estimar

Figura 2

Estimación de la función de supervivencia mediante los tres estimadores descritos, para los datos sobre el tiempo de Complejo Migratorio Motor



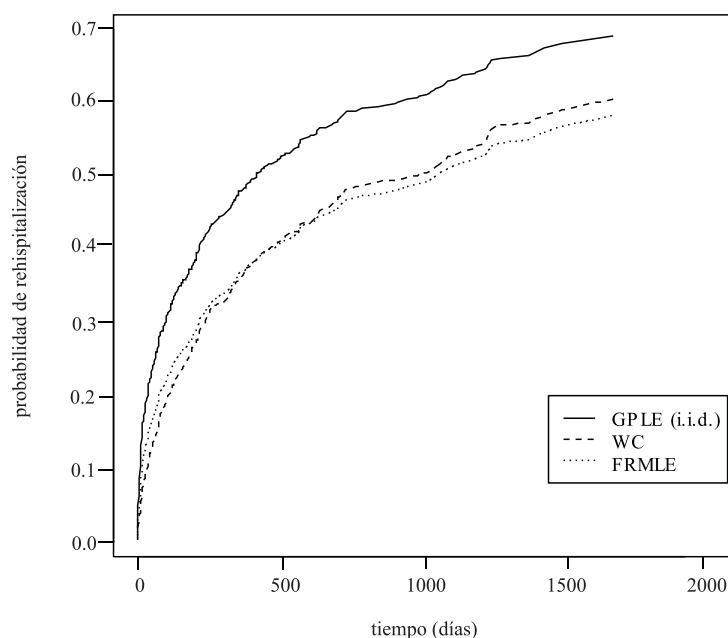
la media de supervivencia con un modelo que asume independencia entre los tiempos de ocurrencias cuando hay correlación puede llevar a conclusiones erróneas. Peña et al. sugieren utilizar un método gráfico para testar dicha hipótesis basándose en los resultados que obtienen en una simulación⁴. La idea es estimar la función de supervivencia utilizando el estimador WC, y los dos propuestos por Peña et al.: el GPLE (modelo i.i.d.) y FRMLE (modelo correlacionado) y en caso que la hipótesis de independencia sea correcta, los tres estimadores deben ser similares. En la figura 2 podemos ver la estimación de la función de supervivencia para estos tres estimadores. En ella se observa la gran coincidencia entre ellos, por lo que se puede concluir que la aserción de Aalen y Husebye es correcta. La estimación del parámetro de fragilidad α es 10,18 lo que nos indica una asociación débil entre los tiempos de interocurrencia (valor de α «grande»). El

resultado obtenido por Aalen y Husebye (1991, tabla 3) bajo el modelo Weibull es $\alpha = 6,85$. En cuanto a los resultados que se obtienen podemos decir que: la media del tiempo de CMM (en minutos) es 104,0 utilizando el estimador WC, 106,0 utilizando GPLE y 108,0 si nos basamos en FRMLE. La estimación obtenida por Aalen y Husebye¹⁹ con el modelo de componentes de la varianza es 106,8 minutos, mientras que bajo el modelo paramétrico de Weibull es de 107,7 minutos. Así, para estos datos, podemos concluir que la estimación de la media del tiempo de CMM no se ve muy afectada por el método de análisis.

Ejemplo 2: El segundo ejemplo trata sobre el tiempo de rehospitalización en pacientes diagnosticados de cáncer colorectal que han sido intervenidos quirúrgicamente. El principal objetivo de los autores fue analizar si existían diferencias en la probabi-

Figura 3

Estimación de la función de supervivencia mediante los tres estimadores descritos, para los datos del ejemplo de rehospitalizaciones y cáncer colorectal



lidad de reingreso en pacientes con cáncer colorectal tras ser intervenidos quirúrgicamente en función de variables clínicas y socio-demográficas. Los autores analizaron numerosas variables, pero para ilustrar este ejemplo utilizaremos dos de ellas: el estadio tumoral (Dukes A-B, C y D) y la edad (<60 , $60-74$, ≥ 75). Los autores están interesados en calcular la probabilidad de reingreso por lo que las gráficas representan la función de distribución, F , en vez de la función de supervivencia, S . En la figura 3 puede observarse la estimación de la función de distribución de los tiempos de rehospitalización utilizando los tres estimadores descritos con anterioridad. Los resultados difieren claramente de los obtenidos en el ejemplo 1. Se puede observar como la estimación WC y FRMLE no coinciden con la de GPLE (modelo i.i.d.). Por ello, y siguiendo las reco-

mendaciones de Peña et al. (2001), podemos concluir que los tiempos de rehospitalizaciones están claramente correlacionados. Así, los investigadores estiman la probabilidad de rehospitalización utilizando el estimador FRMLE. En la tabla 1 puede observarse la probabilidad a 1 y 3 años y el tiempo mediano de reingreso. Observamos como la probabilidad de reingresar (tanto al año como a tres) es mucho mayor a medida que el estadio tumoral es más avanzado. Por ejemplo la probabilidad de reingreso al año para los pacientes con estadios A y B es de 0,26 frente al 0,64 para los pacientes con estadio D. En cuanto a la edad no se observan grandes diferencias. En referencia al tiempo mediano de reingreso, los pacientes con un estadio D tienen una mediana de reingreso de 199 días frente a los 2175 de los pacientes con estadios A y B. Para contestar a la pregunta de si

Tabla 1

Estimación de la probabilidad de reingreso (mediante el estimador FRMLE) y de la mediana de supervivencia (estimadores FRMLE y GPLE) para los datos de rehospitalizaciones y cáncer colorrectal

	Probabilidad de rehospitalización		Mediana de rehospitalización (días)	
	Al año	A los tres años	Modelo de fragilidad (FRMLE)	Modelo de independencia (GPLE)
Edad				
<60	0,39	0,54	799	398
60-74	0,36	0,48	1230	416
≥75	0,33	0,49	1188	524
Estadío Dukes				
A-B	0,26	0,41	2175	1157
C	0,38	0,50	1073	398
D	0,64	0,89	199	107

las diferencias que observamos son estadísticamente significativas se pueden utilizar técnicas de remuestreo²⁰.

Pero ¿hubieran diferido los resultados si se hubiera considerado que las rehospitalizaciones para cada individuo se producen de manera independiente? Fijándonos, por ejemplo, en la mediana de reingreso, obtendríamos que la misma para los pacientes con estadios tumorales A-B sería de 1.157 días, cuando en realidad es de 2.175 días. Esta diferencia, que también se observa en el resto de variables (ver tabla 1, columnas finales), son debidas a la incorrecta elección del modelo que subyace en los datos. El hecho de no tener en cuenta la correlación existente en los tiempos de rehospitalización hace que los resultados difieran ostensiblemente.

COMENTARIOS

Una estrategia comúnmente empleada en este tipo de estudios es analizar tan sólo el tiempo hasta el primer evento y considerar éste como representativo de lo que ocurrirá con posterioridad en cada individuo^{21,22}. Sin embargo esta estrategia tan sencilla, pues se pueden emplear metodología más estándar

como el estimador de Kaplan-Meier, puede llegar a producir resultados muy sesgados²³. Es por ello que se recomienda utilizar metodologías más sofisticadas, como las que se han descrito en este trabajo, pero que tienen en cuenta la verdadera naturaleza de los datos.

También se ha podido comprobar cómo la elección del modelo que genera los tiempos de interocurrencia resulta esencial. Para determinarlo correctamente se propone utilizar un test gráfico que nos ayude a determinar si el modelo de independencia o el de fragilidad es el correcto. Sin embargo si este resultado no es demasiado concluyente una buena solución es utilizar el estimador propuesto por Wang y Chang² que contempla ambas situaciones.

En definitiva, para evitar posibles sesgos es recomendable utilizar las técnicas estadísticas correctas que se ajusten a la naturaleza de los datos. En el caso particular del análisis de supervivencia con eventos recurrentes también es necesario realizar una buena elección del modelo con el que se está trabajando. Para todo ello podemos ayudarnos del software estadístico que disponemos, el cual nos permite obtener los resultados y las con-

clusiones correctas para el problema al que nos estemos enfrentando.

AGRADECIMIENTOS

A Toni Berenguer, por leer una versión preliminar del manuscrito y hacer comentarios que han ayudado a mejorar el mismo. A Guillem Pera y Antonio Agudo por sus consejos referentes a una versión final del escrito.

Anexo 1

El estimador PSH se define utilizando la notación de dos procesos contadores doblemente indexados, $N(s, t)$ e $Y(s, t)$ que se representan como:

$$K_i(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq s\};$$

$$N(s, t) = \sum_{i=1}^n \sum_{j=1}^{K_i(s)} I\{T_{ij} = t\};$$

$$Y(s, t) = \sum_{i=1}^n \left[\sum_{j=1}^{K_i(s-)} I\{T_{ij} \geq t\} + I\{\min(s, \tau_i) - S_{iK_i(s-)} \geq t\} \right]$$

De la misma forma para definir el estimador WC necesitamos definir dos nuevos procesos d^* y K^* de la siguiente forma. Para el individuo i -ésimo, definimos

$$K_i^* = I\{K_i = 0\} + K_i I\{K_i > 0\}$$

entonces

$$d_i^* = \sum_{j=1}^n \frac{1}{K_i^*} \sum_{j=1}^{K_i} I\{T_{ij} = t\};$$

$$R^* = \sum_{i=1}^n \frac{1}{K_i^*} \left[\sum_{j=1}^{K_i} I\{T_{ij} \geq t\} + I\{\tau_i - S_{iK_i} \geq t\} I\{K_i = 0\} \right]$$

Anexo 2

A continuación se describen las instrucciones para obtener las gráficas 1 y 2 correspondientes a los ejemplos del trabajo. Dichas instrucciones deben ejecutarse desde la ventana de comandos del R previa instalación del paquete `survrec`. Esta instalación puede hacerse de manera sencilla utilizando la barra de menús que dispone el R.

Ejemplo 1

```
library(survrec)

data(MMC)

fit<-survfitr(Surv(id,time,event)~1,data =MMC,type="pena")

plot(fit,xlab="tiempo (minutos)")

fit<-survfitr(Surv(id,time,event)~1,data =MMC,type="wan")

lines(fit,lty=2)

fit<-survfitr(Surv(id,time,event)~1,data =MMC,type="MLE")

lines(fit,lty=3)

legend(175,0.9,c("GPLE (i.i.d.)", "WC", "FRMLE"),lty=c(1,2,3),cex=0.8)
```

Ejemplo 2

Los datos referentes a este ejemplo (co- lon) pueden ser pedidos al primer autor.

```
library(survrec)

fit<-survfitr(Surv(HC,tiempo,event)~1, data=colon,type="p")

plot(fit,xlab="tiempo (días)",ylab="probabilidad de rehospitalización", prob= TRUE,type="l")

fit<-survfitr(Surv(HC,tiempo,event)~1, data=colon,type="w")

lines(fit,prob=TRUE,lty=2)

fit<-survfitr(Surv(HC,tiempo,event)~1, data=colon,type="M")

lines(fit,prob=TRUE,lty=3)

legend(1500,0.2,c("GPLE (i.i.d.)", "WC", "FRMLE"),lty=c(1,2,3),cex=0.8)
```

BIBLIOGRAFÍA

1. Cox DR. Regression models and life tables. *J R Stat Soc Ser B* 1972;34:187-220.
2. Wang M, Chang S. Nonparametric estimation of a recurrent survival function. *J Am Stat Assoc* 1999;94:146-153.
3. Lin D, Sun W, Ying Z. Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika* 1999;86:59-70.
4. Peña E, Strawderman R, Hollander M. Nonparametric estimation with recurrent event data. *J Am Stat Assoc* 2001;96:1299-1315.
5. Gill R. Testing with replacement and the product-limit estimator. *The Annals of Statistics* 1981;9:853-860.
6. Vardi Y. Nonparametric estimation in the presence of length bias. *The Annals of Statistics* 1982;10:616-620.
7. Vardi Y. Nonparametric estimation in renewal processes. *The Annals of Statistics* 1982;10: 772-785.
8. McClean S, Devine CA. nonparametric maximum likelihood estimator for incomplete renewal data. *Biometrika* 1995;85:605-618.
9. Soon G, Woodroffe M. Nonparametric estimation and consistency for renewal processes. *Journal of Statistical Planning and Inference* 1996; 53:171-195.
10. Andersen P, Gill R. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982;10:1100-1120.
11. Wei L, Lin D, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 1989;84:1065-1073.
12. Prentice R, Williams B, Peterson A. On the regression analysis of multivariate failure time data. *Biometrika* 1981;68:373-379.
13. Barceló MA. Modelos marginales y condicionales en el análisis de supervivencia multivariante. *Gac Sanit* 2003;16(Supl 2):59-68.
14. Andersen PK, Borgan O, Gill R, Keiding N. Statistical models based on counting process. New York: Springer-Verlag; 1993.
15. Murphy S. Asymptotic theory for the frailty model. *The Annals of Statistics* 1995;23:182-198.
16. Therneau TM, Grambsch PM. Modelling survival data: extending the Cox model. New York: Springer-Verlag; 2000.
17. Sellke T. Weak convergence of the Aalen estimator for a censored renewal process. En *Statistical Decision Theory and Related Topics IV*. S Gupta y J. Berger eds 2. New York: Springer-Verlag;1988.p. 183-94.
18. González JR, Peña EA, Strawderman R. The Survrec Package. The Comprehensive R Archive Network; 2002. Disponible en: <http://cran.r-project.org>
19. Aalen O, Husebye E. Statistical analysis of repeated events forming renewal processes. *Stat Med*, 1991;10:1227-40.
20. González JR, Peña EA. Bootstrapping median survival with recurrent event data. IX Conferencia Española de Biometría; A Coruña; 2003 May 28-30.
21. Hortobagyi GN, Theriault RL, Lipton A, Porter L, Blayney D, Sinoff C, et al. Long-term prevention of skeletal complications of metastatic breast cancer with pamidronate. Protocol 19 Aredia Breast Cancer Study Group. *J Clin Oncol* 1998;16:2038-44.
22. Glynn RJ, Buring JE. Ways of measuring rates of recurrent events. *BMJ*, 1996;312:364-7.
23. Cook RJ, Major P. Methodology for treatment evaluation in patients with cancer metastatic to bone. *J Natl Cancer Inst* 2001;93:534-8.