Modelos de eventos recurrentes aplicados a la industria de producción de aluminio

Recurrent events models applied to the aluminium production industry

Rafael E. Borges[1*], Marianela Luzardo Briceño[1†],

[1]Universidad de Los Andes, Mérida, Venezuela

Resumen

En este trabajo se presenta una revisión actualizada de los principales métodos para analizar datos de eventos recurrentes en la industria. Se presenta además una aplicación en donde se analizan los datos de una selección de los tiempos de vida de 900 celdas de reducción electrolítica de los tres complejos, conformados por líneas de producción, de la planta estatal venezolana de producción de aluminio CVG-Venalum bajo la óptica de los modelos de eventos recurrentes, haciendo énfasis en los sistemas reparables de acuerdo a la naturaleza del proceso. Los datos analizados corresponden al período que va de Enero de 1998 a Junio de 2004.

Palabras clave: Confiabilidad, Análisis de Sobrevivencia, Modelos de Eventos Recurrentes, Aluminio, Producción.

Abstract

In this work, an updated review of the main methods for analyzing recurrent event data in industry is presented. Also, it is presented an application in where the data of a selection of the lifetimes of a 900 cells of electrolytic of the three complexes, conformed by production lines of the state aluminum production plant of Venezuela, CVG-Venalum. This analysis was performed by the methods of the recurrent event models, emphasizing on the repairable systems, because the kind of the process. The analyzed data corresponds to the period between January 1998 and June 2004.

Keywords: Reliability, Survival Analysis, Recurrent Events Models, Aluminum, Production.

1. Introducción

En el estudio del tiempo de seguimiento hasta la ocurrencia de un evento, tópico comúnmente conocido como análisis de supervivencia o confiabilidad, pudiera darse el caso de que a un individuo se le puede observar el evento de interés en más de una ocasión. Al análisis donde la a cada uno de los individuos se les puede observar en más de una ocación el evento de interés se le conoce como análisis de eventos recurrentes.

Algunos ejemplos de datos de eventos recurrentes se presentan, por ejemplo: al estudiar los multiples ataques de epilepsia, la reaparición de ciertos tipos de tumores, los tiempos de duración de máquinas luego de ser sometidas a reparaciones como carros, turbinas, etc.

Existen varias estrategias de modelamiento para este tipo de datos, algunas de ellas están basadas en los procesos puntuales (point processes), otras basadas en procesos de intesidad (intensity processes), otras basadas en modelos semimarkovianos de cambios de estado, otras basadas en modelos de vulnerabilidad compartidos y finalmente, para el caso de unidades reparables, existen una gama de modelos no paramétricos.

En esta conferencia, presentaremos algunos de estos modelos, complementándose con un ejemplo de datos de tiempos de duración de 50 celdas de reducción electrolítica de los tres complejos, conformados por líneas

[*]Profesor. Email:borgesr@ula.ve
[†]Profesora. Email:nela@ula.ve

de producción, de la planta estatal venezolana de producción de aluminio CVG-Venalum, cuya totalidad de los datos fue analizada bajo la óptica de los modelos univariados por Atuve (Altuve 2005).

Recientemente, aspectos relacionados con los modelos de eventos recurrentes han sido abordados de manera exclusiva o a través de cápitulos en libros de análisis de supervivencia y confiabilidad (Nelson 2003), (Meeker & Escobar 1998), (Kalbfleisch & Prentice 2002), (Hougaard 2000) y en algunos artículos de revisión (Peña & Hollander 2004), (Cai & Schaubel 2004), (Hollander & Sethuraman 2004).

2. Modelos de eventos recurrentes

2.1. Modelos basados en procesos puntuales

Estos modelos son ampliamente desarrollados en los textos de Nelson (Nelson 2003) y el de Meeker y Escobar (Meeker & Escobar 1998).

Estos modelos están basados en la función acumulada media (mean cumulative function o MCF) que es una función no decreciente, cuyo algoritmo de cálculo puede verse en el libro de Meeker y Escobar, página 397 (Meeker & Escobar 1998), en algunas herramientas gráficas y algunas familias paramétricas.

Existen varios modelos basados en procesos puntuales, como: El modelos de proceso de Poisson, el modelo del proceso de Poisson no Homogeneo, el modelos de proceso de renovación, los Modelos con covariables y otros modelos

El modelo de proceso de Poisson se caracteriza por tener una MCF lineal λt y una tasa media de recurrencia λ.

El modelo de proceso de Poisson no homogeneo es una generalización del proceso de Poisson con MCF arbitraria.

Debido a lo anterior, una herramienta muy util para la identificación de estos procesos es un gráfico de la MCF versus el tiempo.

El modelo de renovación se caracteriza por que al fallar una unidad, ésta es reemplazada por una nueva y la MCF depende de la función de distribución acumulada.

En los casos anteriores no puede estudiarse el efecto de las covariables, si esto es de interés pueden utilizarse modelos donde la tasa de recurrencia es expresada en función de las covariables. (Therneau & Grambsch 2000)

Entre los otros modelos se encuentran: los modelos markovianos, los modelos de garantías y los modelos de datos longitudinales.

Estos modelos son fáciles de implementar en herramientas estadísticas como R (R Development Core Team 2006), S-PLUS (Insightful Corporation 2001) y SAS (SAS Institute Inc. 2004). Una herramienta de mucha utilidad es la libreria SPLIDA (Meeker & Escobar 2004).

Estos modelos también resultan útiles para comparar dos o más MCF.

2.2. Modelos basados en procesos de intensidad

Estos modelos se derivan de los procesos de intensidad, definidos en el enfoque de análisis de supervivencia basado en los procesos de conteo (Andersen, Borgan, Gill, & Keiding 1993), (Fleming 1993).

Dentro de estos modelos existen cuatro grandes enfoques de modelos semiparamétricos (Cai & Schaubel 2004), (Kalbfleisch & Prentice 2002), estos enfoques son:

- Modelos de regresión condicional.
- Modelos de riesgos marginales.
- Modelos intermedios entre intensidad condicional y riesgos marginales.
- Modelos marginales de medias y/o tasas.

En las siguiente subsecciones procederemos a describir brevemente cada uno de estos modelos.

Todos estos modelos pueden ejecutarse mediante simples variantes a las rutinas ideadas para ajustar los modelos de Cox (Therneau & Grambsch 2000) o mediante simples manipulaciones a los datos (Hosmer & Lemeshow 1999).

2.2.1. Modelos de regresión condicional

Modelo de Andersen y Gill:

Dentro de los modelos de regresión condicionales, el más popular en un modelo basado en procesos de conteo propuesto por Andersen y Gill (Andersen & Gill 1982),

En este modelo, la función de riesgo para el i-ésimo individuo en el tiempo t toma la expresión:

$$\lambda_i(t) = \lambda_0(t)\, e^{\beta_0^T \mathbf{Z}_i(t)}$$

Donde $\lambda_0(t)$ es la función de riesgo base, β_0 es un vector de parámetros desconocidos y $\mathbf{Z}_i(t)$ es una matriz de covariables.

Este modelo es una extensión del modelo de Cox (Cox 1972) y posee dos características esenciales:

- Toda la influencia de los eventos previos sobre las recurrencias futuras, si la hay, son el resultado de covariables dependientes del tiempo.
- Las covariables tienen un efecto multiplicativo en la tasa instantánea del proceso de conteo.

La primera de las características, hace que este modelo sea también conocido como uno de los modelos de reparaciones perfectas, ya que las recurrencias no dependen de la ocurrencias previas.

La segunda de las características, hace que este modelo también sea conocido como el modelo de intensidad multiplicativa.

Modelo de Prentice, William y Peterson:

Un segundo modelo de regresión condicional utilizado frecuentemente es el modelo de Prentice, William y Peterson (Prentice, Williams, & Peterson 1981), en el cual se asume que la intensidad del evento para un individuo o unidad que recién acaba de experimentar la ocurrencia de un evento es idéntica a la intensidad justo antes de la ocurrencia del evento.

En este modelo, la función de intensidad para el individuo i en el tiempo t para la k-ésima ocurrencia viene dada por:

$$\lambda_{ik}(t) = Y_{ik}(t)\, \lambda_{0k}(t)\, e^{\beta_0^T \mathbf{Z}_i(t)} = Y_{ik}(t)\, \lambda_{0k}(t - T_{i,k-1})\, e^{\beta_0^T \mathbf{Z}_i(t)}$$

En este sentido este modelo también es conocido, como el modelo de mínima reparación. Este modelo es un tipo especial de modelo estratificado de Cox que permite que la forma de la función de riesgo dependa de el número de eventos precedentes y posiblemente de otras características.

Modelo para riesgos de tiempos de recurrencia:

Una tercera clase de modelos de regresión condicional es el propuesto por Chang y Wang (Chang & Wang 1999), modelo en el cual la función de riesgo toma la expresión:

$$\lambda_{ij}(t) = \lambda_{0j}(t - T_{i,j-1})\, e^{\beta_0^T \mathbf{Z}_{i1}(t) + \gamma_0^T \mathbf{Z}_{i2}(t)}$$

Que se reduce al modelo de Pentice, Williams y Peterson cuando $\beta_0 = \mathbf{O}$.

Existen otras variantes que pueden ser consultadas en el trabajo de Cai y Schaubel (Cai & Schaubel 2004), donde también pueden verse consideraciones acerca de la estimación.

El modelo de Chang y Wang (Chang & Wang 1999) se encuentra disponible en el paquete survrec (González, Peña, & Strawderman 2006) del lenguaje R (R Development Core Team 2006)

2.2.2. Modelos de riesgo marginal de Wei, Lin y Weissfeld

Otra clase de modelos de regresión, son los conocidos como modelos marginales, del cual existen diferentes variantes, siendo la más utilizada la propuesta por Wei, Lin y Weissfeld (Wei, Lin, & Weissfeld 1989).

Este modelo es también una generalización del modelo de Cox en donde la función de riesgo toma la expresión:

$$\lambda_{ik}(t) = \lambda_{0k}(t)\, e^{\beta_k^T \mathbf{Z}_{ik}(t)}$$

Estos son modelos que tienen la ventaja de que muchas de las facilidades del modelo de Cox pueden ser incorporadas de forma sencilla, pero tiene la desventaja de que las dependencias entre los tiempos de interocurrencia de los eventos no son tomados en cuenta de manera explícita.

2.2.3. Modelos de tasa de Pepe y Cai

Otro modelo, también basado en procesos de conteos es el llamado modelo de tasas y medias o modelo de Pepe y Cai (Pepe & Cai 1993).

Este es una variante del modelo de Andersen y Gill en donde se asume que las covariables tienen un efecto multiplicativo sobre la media y la función de las tasas del proceso de conteo.

En este modelo la tasa de ocurrencia del k-ésimo evento sobre los individuos en riesgo en el tiempo t, viene dada por:

$$r_{ik}(t) = \lim_{x \to o} \tfrac{1}{\delta} P\left[t < T_{i,k} \leq t + \delta | T_{i,k} \geq t, T_{i,k+1} < t\right]$$

2.2.4. Modelos de medias/tasas marginales

Existen por dos modelos de medias/tasas marginales presentados en la literatura.

El primero de los modelos fue propuesto por Lawless y Nadeau (**?**), autores que proponen dos modelos uno semiparamétrico y otro paramétrico definidos mediante:

$$E\left[dN_i(s)\right] = m_0(s)\, g\left(s : \beta_0, \mathbf{Z}_i(s)\right)$$

y

$$E\left[dN_i(s)\right] = m_0(s : \alpha)\, g\left(s : \beta_0, \mathbf{Z}_i(s)\right)$$

respectivamente.

La inferencia de estos modelos para tiempos continuos y para funciones de Cox fue desarrollada posteriormente por Lin et. al. (Lin, Yang, & Ying 2000)

2.3. Clase de Modelos de Peña y Hollander

Peña y Hollander (Peña & Hollander 2004) han propuesto recientemente una nueva clase de modelos que engloban a los anteriores y que son los únicos modelos de eventos recurrentes con capacidad de considerar los efectos de intervenciones, los efectos de acumulaciones de eventos y el efecto de variables concomitantes. Estas características hacen a éste, un modelo atractivo para ser considerado en el modelamiento de eventos recurrentes.

El modelo de Peña y Hollander propone que el proceso de tasa de intensidad para tiempos continuos puede ser expresado mediante:

$$\lambda(s|\mathbf{X}) = \lambda_0(\varepsilon(s))\, \rho\left[N(s-)\right] \psi\left(\beta^T \mathbf{X}\right)$$

donde $\lambda_0(\cdot)$ es la función de tasa de riesgo base, $\rho[\cdot]$ es una función no creciente de $N = \{0, 1, 2, \cdots\}$ en $\Re_+$, $\psi(\cdot)$ es la función de enlace,

$$\beta$$

es un vector de parámetros desconocidos, $\mathbf{X} = \{X_1, \cdots, X_q\}^T$ es el vector de covariables, usualmente dependiente del tiempo y $\{\varepsilon(s) : s \in [0, t]\}$ es un proceso observable y predecible que satisface ciertas condiciones de convergencia, a este proceso se le suele denominar edad efectiva de la unidad (o individuo).

Bajo ciertas condiciones, este modelos puede transformarse en el modelo de Prentice, William y Peterson (Prentice et al. 1981) o en el modelo markoviano propuesto por Gail, Santner y Brown (Gail, & Brown 1980) y en algunas de sus generalizaciones.

Otros modelos que pueden ajustarse mediante el modelo de Peña y Hollander (Peña & Hollander 2004) son algunos modelos no paramétricos para modelos de reparación, como por ejemplo: el modelo de reaparación perfecta, el modelo de mínima reparación, el modelo de Brown y Proschan (Brown & Proschan 1982) el modelo de Block, Borges y Savits (Block, Borges, & Savits 1982),el modelo de Kijima (Kijima 1989), el modelo de Dorado, Hollander y Sethuraman (Dorado, Hollander, & Sethuraman 1997) y el modelo de Last y Szekli (Last & Szekli 1998). Una revisión de estos y otros modelos no paramétricos puede verse en el trabajo de Hollander y Sethuraman (Hollander & Sethuraman 2004).

Los recursos para modelar la clase de modelos de Peña y Hollander está disponible en software, a través del paquete gcmrec (González, Slate, & Peña 2005) del lenguaje R (R Development Core Team 2006).

2.4. Otros Modelos

Existen otros modelos implementados en software, específicamente en el Leguaje R (R Development Core Team 2006), estos modelos son: El modelo de vulnerabilidad Gamma y el modelo de las medias del estimador del producto límite generalizado (Peña, Strawderman, & Hollander 2001), disponibles en el paquete survrec (González et al. 2006) del lenguaje R (R Development Core Team 2006).

Otras estrategias de modelamiento de datos de eventos recurrentes son: Uso de los modelos de cambios de estados mediante cadenas semimarkovianas, uso de modelos de vulnerabilidad compartidos y uso de otros modelos marginales y de cópulas. Algunos de estas estrategias están implementadas en software de forma incompleta. Para una revisión exhaustiva puede consultar el texto de Hougaard (Hougaard 2000).

3. Ejemplo

En esta sección trabajaremos con una selección de 50 de las 900 celdas de reducción electrolítica (todas de la linea 1) de la planta estatal venezolana de producción de aluminio CVG-Venalum (Altuve 2005).

La figura 1 muestra el Gráfico de los eventos recurrentes, en este caso fue generado utilizando la libreria SPLIDA (Meeker & Escobar 2004) del S-PLUS (Insightful Corporation 2001). Un gráfico similar puede obtenerse con el paquete gcmrec (González et al. 2005) del lenguaje R (R Development Core Team 2006).

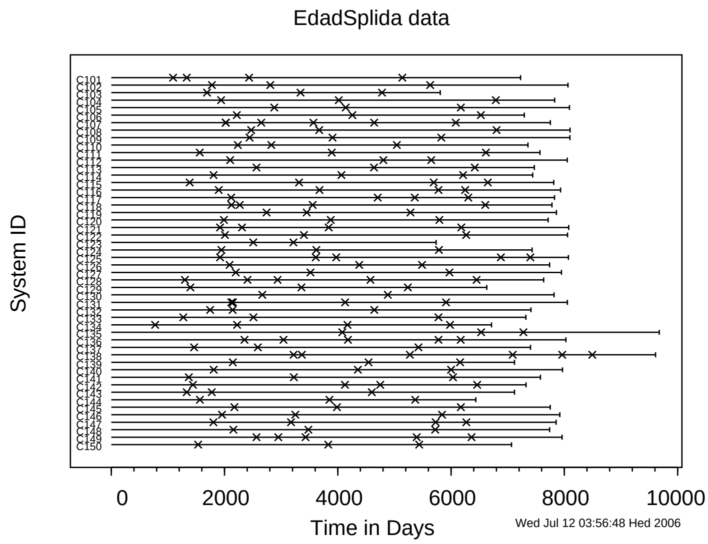

Figura 1: Gráfico de los eventos recurrentes para cada una de las celdas

La figura 2 muestra el gráfico del conjunto de individuos en riesgo.

La figura 3 muestra la función de media acumulada (MCF). En este gráfico puede observarse un patrón de razón constante, pareciendo indicar que pudiera modelarse a través de un proceso de Poisson homogeneo.

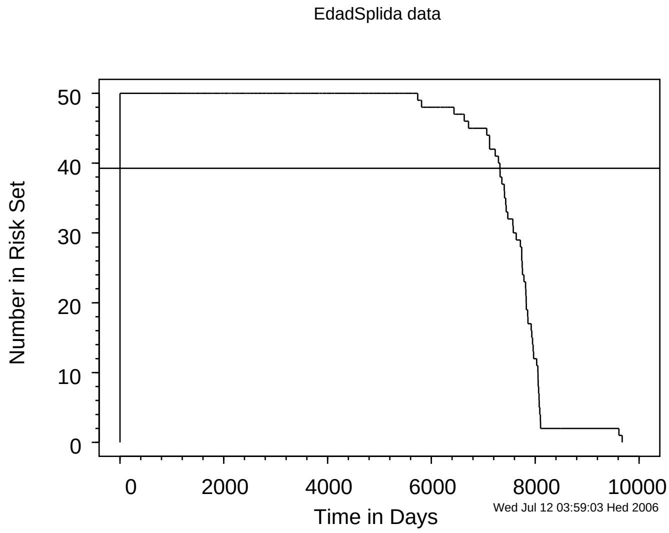

Figura 2: Gráfico del conjunto de individuos en riesgo

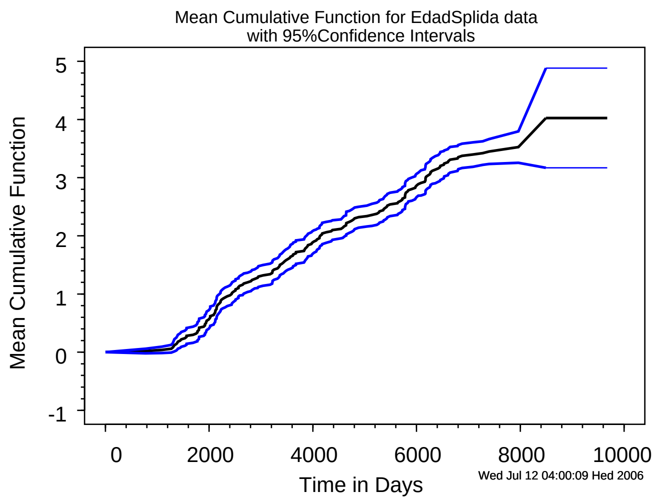

Figura 3: Gráfico de la función de media acumulada (MCF)

Las figuras 4 y 5 muestra los ajustes de dos procesos de Poisson no homogeneos, considerando una regla de potencia y un modelo log-lineal, respectivamente. Puee observarse un mejor ajuste con el modelo log-lineal.

La figura 6 muestra los ajustes de los modelos de vulnerabilidad, de Peña, Strawderman y Hollander y de Wang y Chang, observandose que los dos primero parecieran superponerse mientras que el tercero se ubica sistemáticamente por debajo.

Otros resultados,incluyendo ajustes de la clase de modelos de Peña y Hollander, serán mostrados en la conferencia.

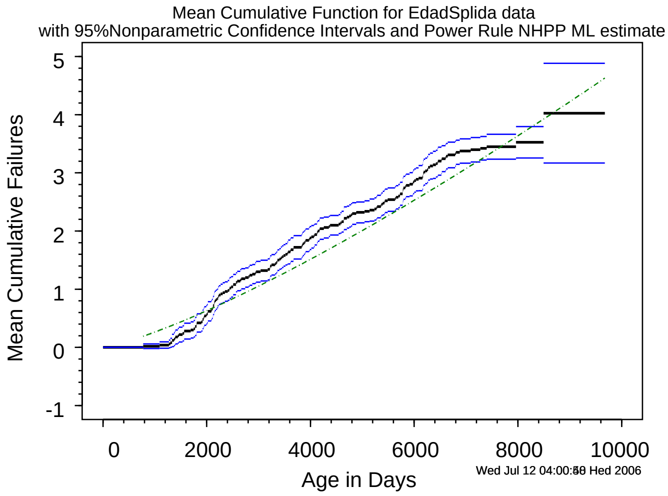

Figura 4: Gráfico del ajuste de un proceso de Poisson no homogeneo utilizando la regla de potencia

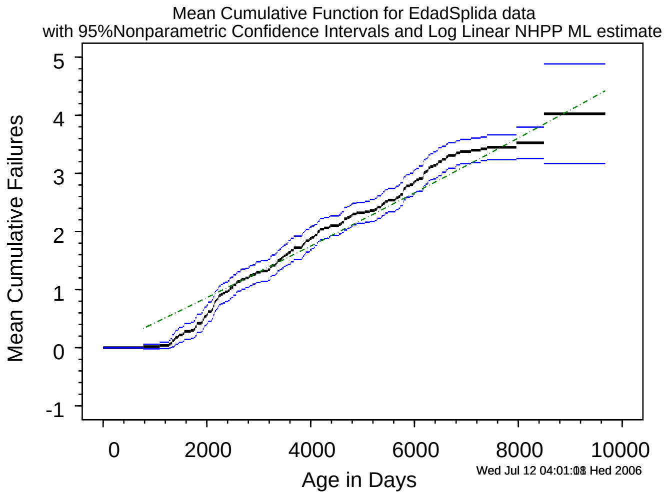

Figura 5: Gráfico del ajuste de un proceso de Poisson no homogeneo utilizando un modelo log-lineal

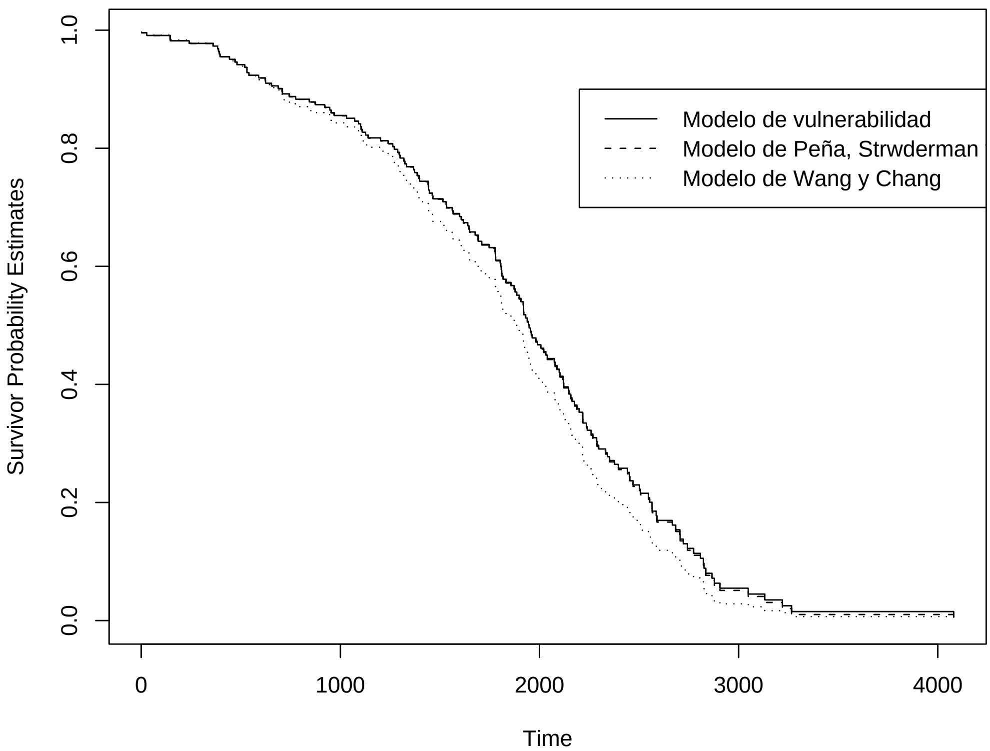

Figura 6: Ajustes de los modelos de eventos recurrentes incluidos en el paquete survrec

Referencias

Altuve, L. (2005), Estimación de vida útil de las celdas de reducción electrolítica. c.v.g. venalum. (enero 1.998 - junio 2.004), Trabajo Especial de Grado, Universidad de los Andes, Facultad de Ciencias Económicas y Sociales. Escuela de Estadística. Mérida. Venezuela.

Andersen, P., Borgan, ., Gill, R., & Keiding, N. (1993), *Statistical models based on counting processes*, Springer, New York.

Andersen, P. & Gill, R. (1982), 'Coxs regression model for counting processes: A large sample study', *The Annals of Statistics* **10**, 1100–1120.

Block, H., Borges, W., & Savits, T. (1982), 'Age-dependent minimal repair', *journal of Applied Probability* **20**, 370–385.

Brown, M. & Proschan, F. (1982), 'Imperfect repair', *journal of Applied Probability* **20**, 851–859.

Cai, J. & Schaubel, D. (2004), Analysys of recurrent event data, *in* N. Balakrishnan & C. Rao, eds, 'Advances in survival analysis, Handbook of statistics 23', Elsevier, North Holland.

Chang, S. & Wang, M. (1999), 'Conditional regression analysis for recurrent time data', *Journal of the American Statistical Association* **94**, 1221–1230.

Cox, D. (1972), 'Regression models and life tables (with discussion)', *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Dorado, C., Hollander, M., & Sethuraman, J. (1997), 'Nonparametric estimation for general repair model', *The Annals of Statistics* **25**, 1140–1160.

Fleming, T.R. y Harrington, D. (1993), *Counting processes and survival analysis*, John Wiley and Sons, Inc, New York.

Gail, M.and Santner, T., & Brown (1980), 'An analysis of comparative carciogenesis experiments based on multiple time to tumor', *Biometrics* **36**, 255–266.

González, J. R., Peña, E. A., & Strawderman, R. L. (2006), *survrec: Survival analysis for recurrent event data.* R package version 1.1-5.
*http://www.r-project.org

González, J. R., Slate, E. H., & Peña, E. A. (2005), *gcmrec: General class of models for recurrent event data.* R package version 0.9-2.
*http://www.r-project.org

Hollander, M. & Sethuraman, J. (2004), Nonparametric methods for repair models, *in* N. Balakrishnan & C. Rao, eds, 'Advances in survival analysis, Handbook of statistics 23', Elsevier, North Holland.

Hosmer, D. & Lemeshow, S. (1999), *Applied survival analysis: Regression modeling of time to event data*, John Wiley and Sons, Inc, New York.

Hougaard, P. (2000), *Analysis of multivariate survival data*, Springer, New York.

Insightful Corporation (2001), *S-PLUS 6 for Windows Guide to Statistics, Volume 2*, Insightful Corporation, Seattle, WA.
*http://www.insightful.com

Kalbfleisch, J. & Prentice, R. (2002), *The statistical analysis of failure time data*, second edn, John Wiley and Sons, Inc, New York.

Kijima, M. (1989), 'Some results for repairable systems with general repair', *Journal of Applied Probaility* **26**, 89–102.

Last, G. & Szekli, R. (1998), 'Asymptotic and monociticity properties of some repairable systems', *Advances in Applied Probability* **30**, 1089–1110.

Lin, D.Y.and Wei, L., Yang, I., & Ying, Z. (2000), 'Semiparametric regression for the mean andrate functions of recurrent events', *Journal of the Royal Statistical Society, Series B* **62**, 711–730.

Meeker, W. & Escobar, L. (1998), *Statistical methods for reliability data*, John Wiley and Sons, Inc, New York.

Meeker, W. & Escobar, L. (2004), *SPLIDA (S-PLUS Life Data Analysis)*, Iowa State University and Louisiana State University.

Nelson, W. (2003), *Recurrent events data analysis for product repairs, disease recurrences, and other applications*, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

Peña, E. & Hollander, M. (2004), Models for recurrent events in reliability and survival analysys, *in* R. Soyer, T. Mazzuchi, & N. Singpurwalla, eds, 'Mathematical reliability: An expository perspective', Kluwer Academic Publishers, Boston.

Peña, E., Strawderman, R., & Hollander, M. (2001), 'Nonparametric estimation with recurrent event data', *Journal of the American Statistical Association* **96**, 1299–1315.

Pepe, M. & Cai, J. (1993), 'Some graphical displays and marginal regression analysis of recurrent failure times and time dependent covariates', *Journal of the American Statistical Association* **88**, 881–820.

Prentice, R., Williams, B., & Peterson, A. (1981), 'On the regression analysis of multivariate failure time data', *Biometrika* **68**, 373–379.

R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

SAS Institute Inc. (2004), *SAS/STAT 9.1 user's guide*, SAS Institute Inc., Cary, N.C.
*http://www.sas.com

Therneau, T. & Grambsch, P. (2000), *Modeling survival data: Extending the Cox Model*, Springer, New York.

Wei, L., Lin, Y., & Weissfeld, L. (1989), 'Regression analysis of multivariate incomplete failure time data by modeling marginal distributions', *Journal of the American Statistical Association* **84**, 1065–1073.