

MODELOS DE ANÁLISIS DE SOBREVIVENCIA MULTIVARIADOS.

Rafael E. Borges P.

Escuela de Estadística,
Universidad de Los Andes,
Mérida 5101, Venezuela.
e-mail: borgesr@ula.ve

Resumen.

En esta conferencia se presenta una introducción a los modelos de análisis de sobrevivencia multivariados. Se discuten los aspectos generales de los datos de sobrevivencia. Se presentan las principales estructuras de los datos de sobrevivencia multivariados: (i) Estructuras de datos paralelos y, (ii) Estructura de datos longitudinales y los seis principales modelos de datos en sobrevivencia multivariados: (i) Modelos de varios individuos, (ii) Modelos de órganos similares, (iii) Modelos de eventos recurrentes, (iv) Modelos de medidas repetidas, (v) Modelos de eventos diferentes y, (vi) Modelos de riesgos en competencia, acompañados de referencias y comentarios de aplicaciones publicadas en la literatura científica. Se discute acerca de los tipos de censura y truncamiento y los objetivos de los estudios de sobrevivencia multivariados, así como también acerca de los principales métodos de análisis de sobrevivencia multivariado y se hacen comentarios acerca de aspectos computacionales y de libros y revistas sobre el tópico.

1. Introducción.

El análisis de supervivencia multivariado es un área poco tratada en libros especializados, la mayoría del desarrollo está disponible en artículos científicos. El único libro dedicado íntegramente al análisis de sobrevivencia multivariado es del de Hougaard (Hougaard, 2000), texto en el cual está basado gran parte de esta conferencia.

Otro texto que presenta un enfoque actualizado es del de Andersen et.al. (1993), que a pesar de dedicar un solo capítulo al tópico, la rigurosidad matemática con que es abordado, lo hace una referencia obligatoria.

Esta conferencia está estructurada en 10 secciones, incluyendo esta breve introducción, en la segunda sección se introducen las dos principales estructuras de los datos de sobrevivencia multivariados, en la tercera sección se presentan los seis principales modelos. En la sección 4, se presentan algunos mecanismos de censura y truncamiento multivariados, seguido de una sección donde se asoman los objetivos y propósitos del análisis de sobrevivencia multivariado. La sexta sección presenta los diversos métodos con los que se puede abordar el análisis de sobrevivencia multivariado, luego se presenta una sección con comentarios acerca de aspectos computacionales y al material bibliográfico sobre el tema, para seguir con una breve sección de agradecimiento, finalmente, se presenta una lista de las fuentes bibliográficas consultadas para la elaboración de esta conferencia.

2. Estructuras de datos de sobrevivencia multivariados.

En el análisis de sobrevivencia multivariado pueden existir dos estructuras de datos básicas: Datos paralelos y datos longitudinales. Teóricamente, todos los datos asociados a los modelos de sobrevivencia multivariados pueden ser considerados como datos longitudinales usando el tiempo calendario como escala del tiempo. Pero el trabajar ciertas situaciones utilizando una estructura de datos paralelos, cuando se puede, tiene las ventajas que los datos paralelos son más sencillos y pueden aplicarse metodologías que sólo son aplicables a este tipo de datos, como lo son: La estimación de medidas de dependencia y el uso de métodos de estimación no paramétrica.

2.1. Datos paralelos.

Se refiere a la situación en que varios individuos o unidades de un sistema son seguidos simultáneamente.

En este tipo de estructura, el conjunto de datos posee una forma matricial en donde los tiempos de respuesta toman la forma T_{ij} , $i = 1, \dots, n$, $j = 1, \dots, k_i$ donde i representa el grupo y j denota un individuo dentro de cada grupo. Existe un conjunto de indicadores de fallas representados mediante D_{ij} , $i = 1, \dots, n$, $j = 1, \dots, k_i$.

Este modelo posee una serie de características adicionales, como son:

En este modelo se asume que existe independencia entre los distintos valores de i y, posiblemente, dependencia entre los tiempos para el mismo valor de i .

Se asume además que el número de tiempos en está predeterminado, este número puede ser fijo(k) o variable (k_i).

Debido a la presencia de datos censurados, el número de fallas o eventos observado para cada grupo ($\sum_i D_{ij}$) no está especificado y el número de eventos es a lo sumo k o k_i , según sea el caso.

No deben existir restricciones de orden como por ejemplo $T_{i1} < T_{i2}$, ya que la existencia de este tipo de restricciones convertirían la estructura de datos en una correspondientes a datos longitudinales.

Por último, en esta estructura se pueden considerar tres tipos de covariables: (i) Comunes para cada todos los tiempos en un grupo (dependientes sólo de las i), (ii) ser de tipo pareadas (dependientes solo de las j) y, (iii) diferentes para cada observación o covariables generales (dependientes tanto para las i como para las j).

2.2. Datos longitudinales.

En esta estructura se asume que se observa un proceso estocástico para cada valor de i y se observan eventos o transiciones sobre el tiempo. En este tipo de datos existe un proceso para cada individuo (persona, componente, etc.). Otro nombre que recibe esta estructura es la del análisis de historias de vida.

Las características más importantes de los datos longitudinales son:

Se asume independencia sobre los i .

Los tiempos de transición son una sucesión creciente de tiempos.

En el caso más sencillo, sólo el último tiempo puede ser censurado y todos los demás tiempos son eventos (tiempos de transición).

El número de tiempos k_i es aleatorio.

Algunos modelos tienen un número máximo posible de eventos otros no.

Los tiempos pueden corresponder a eventos recurrentes similares o a transiciones entre varios estados, en cuyo caso, los datos no incluyen solo los tiempos, sino adicionalmente el nombre de la transición.

En la estructura de datos longitudinales, el proceso i es seguido durante el intervalo $(0, C_i)$ donde C_i es un tiempo fijo o aleatorio y los tiempos de los eventos son de la forma T_{ij} , $i = 1, \dots, n$, $j = 1, \dots, k_i$. Dependiendo de las circunstancias, los eventos pueden considerarse como tiempos de eventos puros usando C_i como el final de una observación o, pueden considerarse los tiempos como tiempos de eventos más una censura en el tiempo C_i . En el primer caso es posible que $k_i = 0$.

Deben establecerse restricciones de orden del tipo $T_{i,j} < T_{i,j+1}$, pudiéndose considerar la igualdad como un caso especial que necesitaría consideraciones especiales.

Si el proceso admite varios eventos posibles, es necesario conocer el estado inicial S_{i0} y los tipos de transiciones observadas, las transiciones pueden ser escritas como S_{ij} , $i = 1, \dots, n$, $j = 1, \dots, k_i$ y pueden ser descritos según los nombres de los estados de transiciones que suelen representarse mediante $1, \dots, m$, reservando el valor 0 a la censura. Si el estado inicial es común para cada todos los individuos es preferible denotar a este primer estado mediante el número 1.

Es posible incluir covariables, que pueden ser constantes o dependientes del tiempo.

El modelo puede ser extendido de tal manera que contemple huecos en los datos, es decir que el proceso puede ser seguido para varios períodos no conectados.

Finalmente, en algunos casos no es posible tener los tiempos exactos, sino el número de eventos en un intervalo de tiempo seleccionado.

3. Modelos de datos en análisis de sobrevivencia multivariado.

3.1. Modelo de varios individuos.

Un caso típico de datos paralelos es la determinación de los tiempos de vida de varios individuos que no pueden ser considerados como independientes debido a que están relacionados de alguna manera. Algunos ejemplos de este modelo de datos son los gemelos, que comparten toda el material genético (monocigóticos o idénticos) o parte del material genético (dicigóticos o fraternales), las parejas de casados seguros de vida de empleados que comparten el mismo ambiente en su sitio de trabajo, experimentos pareados, entre otros. Este modelo puede ser utilizado en estudios multicéntricos.

Uno de los conjuntos de datos más trabajados por Hougaard, es de 8985 gemelos daneses que nacieron entre 1881 y 1930 y que estaban vivos a los 15 años, los cuales fueron seguidos hasta 1980 (Hauge et. al, 1968, Hougaard, Harvald y Holm, 1992a, 1992b)

Un segundo ejemplo es el estudio de la comparación de los tiempos de vida entre 960 niños nacidos entre 1924 y 1926 daneses adoptados y sus padres biológicos y adoptivos (Sørensen et. al., 1988), este conjunto de datos es estudiado también por Nielsen et. al. (1992).

Mantel, Bohidar y Cimenera (1997) presentan un conjunto de datos que consiste en 50 camadas de ratas del sexo masculino y 50 camadas del sexo femenino, en donde en cada camada , una de las ratas fue tratada con una droga y las otras dos fueron usadas como controles y se realizó un seguimiento de 104 semanas hasta observar la aparición de tumores. Este conjunto de datos es reanalizado por Nielsen et. al. (1992).

En un estudio de comparación de tratamientos, Freireich et. al. (1963) comparan el tiempo de remisión de leucemia en 21 pares de pacientes, apareados según el centro y el estatus de remisión, en los cuales a uno de los pacientes se le administró la droga 6-Mercaptopurina y al otro se le administró un placebo. Este conjunto de datos fue reanalizado por Andersen et. al. (1993).

Otro ejemplo abordado en la literatura es el análisis de los datos del Farmingham Heart Study, analizando la información de los individuos dentro de las familias, asumiendo que presentan un efecto genético y de compartimiento de ambientes (Klein, 1992, Nielsen et. al, 1992).

3.2. Modelo de órganos similares.

Otro modelo de datos paralelos es el de órganos similares en el cual se estudian tiempos de falla de órganos similares, como por ejemplo ojos (derecho e izquierdo), riñones, pulmones, dientes, entre otros. Estos datos son poco comunes en medicina pero son muy frecuentes en aplicaciones industriales en donde se comparan las fallas de las componentes de un sistema.

Huster, Brookmeyer y Self (1989) presentan un estudio comparativo de dos tratamientos para la ceguera para 197 pacientes diabéticos, a los cuales se les aplicó de manera aleatorizada un tratamiento de fotocoagulación por laser, dejando el otro ojo como control. Este conjunto de datos fue reanalizado por Liang, Self y Chang (1993).

Otro ejemplo de este tipo de datos es uno referido al llenado de 566 amalgamas para 32 personas realizadas por 6 odontólogos durante 17 años en la cual se estudió el efecto del sexo y la edad en 1987 sobre el riesgo de fractura de las amalgamas (Aalen, Bjertness, y Sønju, 1995)

3.3. Modelo de eventos recurrentes.

Este es un caso interesante y de mucha utilidad de datos longitudinales en individuos que pueden presentar tipos de eventos similares en forma recurrente. Han sido aplicados en problemas de fertilidad femenina, en el estudio de patrones menstruales, en estudios de efectos adversos a los medicamentos, como por ejemplo múltiples episodios de hipoglicemia en pacientes diabéticos y en modelos de apariciones de tumores múltiples en algunos órganos, siendo este último caso en algunas ocasiones difíciles de estudiar, debido a que en algunos caso suele ser difícil la identificación de tumores separados.

Gail, Santner y Brown (1980) presentaron un conjunto de datos de apariciones de tumores mamarios en un grupo de ratas hembras a las cuales se les inyectó un agente cancerígeno al día 0 y fueron tratadas con acetato retinyl por 60 días y luego de comprobar que 48 de las ratas estaban libres de tumores, se continuó el tratamiento en 23 con el tratamiento retinoide y las otras 25 fueron tomadas como grupo control y continuó el tratamiento hasta el día 182, observando el día de aparición de cada tumor.

Thall y Vail (1990) presentan un conjunto de datos para ataques epilépticos, en donde se cuentan el número de episodios en dos grupos de pacientes que recibían medicamentos antiepilépticos, un grupo recibió adicionalmente progabide y otro placebo.

McGilchrist y Aisbet (1991) presentan un análisis de múltiples episodios de infecciones por catéteres en 38 pacientes que utilizaban equipos portátiles de diálisis.

En este mismo sentido, tenemos un conjunto de datos para analizar episodios múltiples de peritonitis en pacientes que acuden al Servicio de diálisis peritoneal del Hospital Clínico Universitario de Caracas. Otro de los problemas que se tienen planteados es el de la construcción de modelos para explicar las múltiples infecciones de malaria causada por el parásito *Plasmodium falciparum* en Tumeremo, Estado Bolívar, Venezuela.

3.4. Modelo de medidas repetidas.

Este tipo de datos provienen de resultados de experimentos diseñado en donde se estudia el mismo tipo de eventos estudiados en un número fijo de tiempos para cada individuos. Los datos obtenidos son similares a los de los eventos recurrentes, sin embargo la estructura de datos es paralela debido a que el número de observaciones del tiempo es fija. Una seria limitación de este modelo de datos es que es necesario que el último tiempo de observación corresponda a una censura.

Un ejemplo interesante es el presentado por Danahy et. al. (1976), en donde estudian el tiempo transcurrido hasta producirse una angina de pecho antes y después de tratamientos con dinitrato isosórbido, nitroglicerina y placebo en 21 pacientes sometidos a 10 ejercicios respiratorios (1 antes del tratamiento, y tres después de cada tratamiento transcurridas 1 hora, 3 horas y 5 horas). Este conjunto de datos fue nuevamente analizado por Pickles et. al. (1994).

3.5. Modelo de eventos diferentes.

En este modelo de datos se tiene una estructura en donde los individuos son seguidos en distintos estados y lo que se modela son los cambios de estados. En este tipo de modelos, resulta de particular interés como los distintos eventos influyen en el curso futuro de los individuos. En este tipo de datos, el número de eventos suele ser bajo. Existen dos modelos típicos como lo son el modelo de discapacidades y el modelo de sacrificios.

Un ejemplo de datos similares a los del modelo de discapacidades es uno en que se estudia la mortalidad de 328 pacientes diabéticos no insulino dependientes de acuerdo a su estado de albuminuria (normoalbuminuria, microalbuminuria y macroalbuminuria) en 1987.

Moertel et. al. (1990) y Fleming (1992) presentan resultados de un ensayo clínico para estudiar el tiempo de recurrencia y muerte por cáncer de colon en 926 pacientes en la etapa C entre 1985 y 1987, los cuales fueron aleatorizados en tres grupos que recibieros solo lemisole, lemisole combinado con flouorouacil o, simplemente fueron observados.

Hougaard (1986) utiliza una base de datos para el seguimiento de una año, haciendo énfasis en las primeras seis semanas, de 1059 pacientes que ingresaron por infarto al miocardio a un hospital.

Un conjunto de datos ampliamente estudiado y presentado en la literatura es el del Stanford Heart transplant program, donde el interés estaba centrado en determinar si los trasplantes de corazón mejoran la sobrevida de los pacientes (Crowley y Hu, 1977, Kalbfleisch y Prentice, 1980, Allison, 1995).

3.6. Modelo de riesgos en competencia.

Los modelos de riesgos en competencia se presentan cuando se efectúa el seguimiento de cada individuo hasta la ocurrencia de solo uno de k posibles eventos. Este modelo es

considerado como un caso particular de datos longitudinales con un número de eventos de a lo sumo uno.

El modelo de riesgos en competencia no es estrictamente un modelo de análisis de sobrevivencia multivariado, aunque textos clásicos de los años 80 lo habían considerado como uno de los casos más emblemáticos de lo que hasta ese entonces se conocía como el análisis de sobrevivencia multivariado (Kalbfleisch y Prentice, 1980, Lawless, 1982, Cox y Oakes, 1984).

Sin embargo, el considerar al modelo de riesgos en competencia como un modelo multivariado tiene sus ventajas. Por un lado se beneficia de la teoría del análisis de sobrevivencia multivariado y adicionalmente, enriquece a los otros modelos cuando lo que se estudia es el tiempo hasta el primer evento.

Crowder (2001) presenta una visión amplia y actualizada de los modelos de riesgos en competencia.

Hoel y Walburg (1972) describen un conjunto de datos de 95 ratones convencionales y 82 ratones libres de gérmenes los cuales fueron expuestos a una radiación de rayos X de 300 rads entre las 5 y semanas de vida. Se consideraron tres posibles causas de muerte: Linfoma tímico, sarcoma de célula reticular y otras causas.

Andersen et.al. (1993) presentan un análisis de 205 pacientes que fueron sometidos a una operación radical para melanoma maligno, considerando a la muerte por melanoma y muerte por otras causas como los dos posibles eventos.

Actualmente y bajo la perspectiva de los modelos de riesgos en competencia, estamos analizando los datos de Borges (2002), tomando a muerte y salida de diálisis por razones médicas como eventos de interés, en 246 paciente de diálisis peritoneal y 140 pacientes de hemodiálisis.

4. Censura y truncamiento en el análisis de sobrevivencia multivariado.

Los mecanismos de censura para los datos de sobrevivencia multivariados son mucho más complejos que para el caso univariado.

Para las estructuras de datos paralelos existen dos tipos de censura: Censura homogénea y censura heterogénea.

La censura homogénea, llamada también censura univariada, se presenta cuando se observa simultáneamente a todo el grupo como un proceso estocástico y el final del estudio o la falla o muerte de cada individuo conduce a una censura simultanea de cada uno de los individuos que son estudiados. Este es el mecanismo de censura típico del modelo de órganos similares.

La censura heterogénea se presenta cuando cada individuo tiene tiempos de censura separados y generalmente son diferentes.

Para el caso de datos longitudinales, lo más común es estudiar cada proceso desde el tiempo 0 hasta el tiempo C_i , que suele ser diferente de proceso a proceso.

El caso de los datos truncados en datos de análisis de sobrevivencia multivariado es aún más complejo y en muchos casos es difícil de medir que para el caso univariado, ya que existen muchas formas de medirlo.

Cuando los eventos del mismo tipo, que es el caso más sencillo, suele especificarse que los individuos se encontraban vivos al comenzar el estudio. Para el caso de datos paralelos suele asumirse que no han ocurrido eventos hasta el momento de comenzar el seguimiento. Para datos longitudinales uno pudiera tener o no tener información acerca del número de eventos que ha experimentado cada individuo al momento de comenzar su seguimiento.

Cuando los eventos son de distinto tipo, la información del truncamiento debe determinarse para cada tipo de eventos.

5. Objetivos del análisis de sobrevivencia multivariado.

Los objetivos del análisis de sobrevivencia multivariado son varios y varios de ellos serán mencionados, sin entrar en detalles, en la presente sección:

Un primer objetivo, que en muchos casos no se puede cumplir es la estimación de la función de sobrevivencia conjunta.

Un segundo propósito es la determinación de los efectos de las covariables comunes de manera general.

Un tercer objetivo es la determinación de los efectos de las covariables pareadas.

El cuarto gran objetivo es la evaluación de la dependencia entre las posibles repuestas de los datos, para tratar de identificar distintos niveles de riesgos.

El quinto objetivo es la predicción, es decir, estimar la probabilidad de ocurrencia de un evento basado en la información recolectada anteriormente o durante el estudio.

El último objetivo considerado es el chequeo del modelo que se hace con la intención de verificar si corresponde a la teoría.

Uno de los aspectos que deben ser considerados a la hora de la formulación de un modelo es la verificación de los supuestos, este paso es necesario para que los modelos sean considerados como válidos.

6. Métodos de análisis de sobrevivencia multivariados.

En esta sección se presenta diversos métodos de análisis de sobrevivencia multivariados. Los métodos que se presentan son los de la estimación no paramétrica de la función de sobrevivencia multivariada, los modelos de regresión conocidos como modelos frágiles, el enfoque basado en distribuciones marginales y cópulas y, el enfoque basado en los modelos multiestados. Una visión introductoria es presentada en Hougaard (1986).

6.1. Estimación no paramétrica de la función de sobrevivencia multivariada.

Una aproximación elegante al problema de estimación en el contexto del análisis de sobrevivencia multivariado es presentado por Dabrowska (1998), la cual es presentada a continuación:

Asumamos que el vector multivariado de interés es de la forma (T, \mathbf{Z}) donde T es un tiempo de falla univariante sujeto a censura y \mathbf{Z} es un vector de covariables no censuradas. En este caso, los métodos de estimación no paramétrica están basados en recobrar la los parámetros de la función de sobrevivencia condicional:

$$S(t|\mathbf{z}) = P(T > t | \mathbf{Z} = \mathbf{z}) = \pi_{[0,1]}(1 - A(du|\mathbf{z})) \quad (1)$$

donde $A(t|\mathbf{z})$ la función de riesgo acumulada de la distribución condicional del tiempo de falla T dada la covariable $\mathbf{Z}=\mathbf{z}$.

Los métodos de estimación se basan en los estimadores de Beran.

El estimador de Nelson y Aalen toma la forma:

$$\hat{A}(t|\mathbf{z}) = \int_{[0,t]} \frac{\sum_{i=1}^n W_i(\mathbf{z}) N_i(du)}{\sum_{i=1}^n W_i(\mathbf{z}) Y_i(u)}$$

donde (N_i, Y_i) , $i = 1, \dots, n$ son los procesos de conteo y de riesgo, $N_i(t) = I(T_i \leq t \wedge \tilde{T}_i)$ y $Y_i(t) = I(T_i \wedge \tilde{T}_i \geq t)$ asociados con la muestra de los n individuos en estudio, donde \tilde{T} es el tiempo de censura dadas las covariables \mathbf{Z} y, $W_i(\mathbf{z})$, $i = 1, \dots, n$ son los pesos, que dependen únicamente de las covariables $\mathbf{Z}_1, \dots, \mathbf{Z}_n$.

El estimador de Kaplan y Meier puede obtenerse mediante:

$$\hat{S}(t|\mathbf{z}) = \pi_{[0,1]} \left(1 - \int_{[0,t]} \frac{\sum_{i=1}^n W_i(\mathbf{z}) N_i(du)}{\sum_{i=1}^n W_i(\mathbf{z}) Y_i(u)} \right)$$

La obtención de estos estimadores es complicado desde el punto de vista matemático. Se han presentado en la literatura varios estimadores para el caso bivariado (Dabrowska, 1998). Una interesante comparación de los estimadores es presentada en Andersen et. al (1993).

Se prefieren los métodos de estimación empíricos basados en procesos puntuales a los basados en procesos de conteo, por no satisfacer estos últimos ciertas condiciones de suficiencia.

6.2. Modelos frágiles.

Los modelos frágiles (frailty models) son modelos de regresión de efectos aleatorios ampliamente utilizados en análisis de sobrevivencia. Fueron introducidos por Vaupel, Manton y Stallard (1979) aunque su popularización se logra desde finales de los años ochenta y principios de los noventa. Una breve introducción es presentada por Hougaard (1998). Una excelente presentación es dada por Andersen et. al (1993). La extensión para el caso multivariado es presentada por Hougaard (2000).

En análisis de sobrevivencia multivariado, los modelos frágiles son usados para inducir la dependencia de los tiempos de falla a través de un efecto aleatorio que recoja otros factores que relacionen los procesos marginales (Dabrowska, 1998).

Los modelos más sencillos se obtienen bajo el supuesto de que los m individuos comparten un factor no observable W , que corresponde a una variable aleatoria no negativa tal que los tiempos de falla $(\mathbf{T}_1, \dots, \mathbf{T}_m)$ condicionados por W y el vector de covariables observadas, $(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$, son independientes de las variables aleatorias con funciones de riesgos:

$$\alpha_i(t|\mathbf{Z}, W) = W_{0_i}(t) \exp(\beta_i^T \mathbf{Z}_i), \quad i = 1, \dots, m$$

Y la función de sobrevivencia condicional de los tiempos de falla $\mathbf{T} = (T_1, \dots, T_m)$ dadas las covariables \mathbf{Z} es:

$$S(t_1, \dots, t_m | \mathbf{Z}) = \int \left[-w \sum_{i=1}^m \exp(\beta_i^T \mathbf{Z}_i) \int_0^{t_i} \alpha_{0_i}(u) du \right] F_W(dw)$$

donde F_W es la distribución de la variable frágil W .

Usualmente, la distribución F_W pertenece a una distribución paramétrica conocida como una Gamma, una Gaussiana inversa o una lognormal. La estimación de estos modelos se basa en el principio de máxima verosimilitud paramétrica y es complicada.

Otros métodos de estimación son los métodos de máxima verosimilitud no paramétricos y los basados en el algoritmo EM.

6.3. Enfoque basado en los modelos de múltiples estados (multi-state).

Este es un enfoque en que se identifica el modelo múltiples estados más adecuado y se modelan las probabilidades de transición entre los estados.

La estimación se hace a través de modelos Markovianos, semimarkovianos, Markovianos ocultos y modelos progresivos. Para una amplia exposición puede consultarse el texto de Hougaard (2000).

6.4. Enfoque basado en las distribuciones marginales y cópulas.

También existen modelos basados en distribuciones marginales y cópulas que son ampliamente discutidos en Hougaard (2000).

Los modelos basados en distribuciones marginales se basan en el conocimiento que se tiene de las distribuciones marginales que no asumen supuestos distribucionales (distribution-free) y en supuestos de dependencia.

En el enfoque basado en cópulas, se estudia la estructura de dependencia, cuando se remueve la influencia de las distribuciones marginales, considerándolas como fijas.

En el contexto de marginales y cópulas también existen enfoques coordinados y combinados (Hougaard, 2000)

7. Aspectos computacionales.

Los datos paralelos suelen ser analizados de manera natural, considerando cada registro para cada valor correspondiente al grupo o, correspondiente a cada individuo y grupo.

Existen herramientas para la estimación no paramétrica bivariada de la función de supervivencia en la página web de librería estadística de la Universidad Carnegie Mellon (<http://lib.stat.cmu.edu>).

Algunas herramientas basadas en modelo frágiles y de modelos de riesgos en competencias están disponibles en SAS, S-PLUS y el lenguaje R (Allison, 1995, Therneau y Grambsch, 2000).

Hougaard (2000) presenta algunas herramientas para el modelaje a través de distribuciones marginales y cópulas para SAS.

8. Comentarios sobre libros y revistas.

El análisis de sobrevivencia multivariado tal y como se enfoca actualmente es un tópico poco incorporado en la mayoría de los libros de análisis de sobrevivencia.

Su tratamiento comienza a ser considerado tímidamente en textos clásicos de los ochenta como los de Kalbfleisch y Prentice (1980), Lawless (1982) y Cox y Oakes (1984), siendo el tratamiento un poco más extenso en el texto de Lawless, los otros dos libros tratan únicamente lo referente a los modelos de riesgos en competencia.

Sin embargo, el enfoque más actualizado es incorporado en libros mucho más recientes. En el texto de Andersen et. al. (1993) se presenta un capítulo con un desarrollo matemático riguroso sobre el tópico. Hougaard (2000) recopila muchos de los modelos y métodos disponibles hasta el momento. Crowder (2001) presenta una revisión detallada de los modelos de riesgos en competencia. Muchos de los aspectos relacionados con el tema son abordados a través de varios artículos en la Enciclopedia de Bioestadística, donde el análisis de sobrevivencia es uno de los temas más ampliamente presentados (Armitage y Colton (editores) (1998). Otros libros importante es el editado por Klein y Goel (1992) y el editado por Lin y Fleming (1997). Una buena visión introductoria es presentada en el texto de Klein y Moeschberger (1997).

También se han publicado libros útiles para resolver aspectos computacionales como el de Allison (1995) y el de Therneau y Grambsch (2000).

En lo que se refiere a las revistas las que más publican artículos sobre el tópico son las de mayor impacto en estadística general como *Biometrics*, *Journal of the American Statistical Association*, *Statistics in Medicine*, *Biometrika*, *Scandinavian Journal of Statistics*, *The Annals of Statistics*, *Journal of the Royal Statistical Society (Series B)* y *Journal of Multivariate Analysis*, entre otras. También se edita desde 1995 una revista especializada en análisis de sobrevivencia, *Lifetime Data Analysis*.

9. Agradecimientos.

Agradezco a la Comisión de Desarrollo Científico, Humanístico y Tecnológico de la Universidad de Los Andes (CDCHT-ULA), a través del proyecto “Comparación de Modelos de Análisis de Supervivencia y Confiabilidad” (Código E-199-02-09-C), fuente de financiamiento que ha permitido la elaboración de esta conferencia.

Agradezco además a los miembros del Comité Organizador del Simposio de Estadística 2004 por su gentil invitación para dictar esta Conferencia.

10. Referencias bibliográficas.

- [1] Aalen, O.O., Bjertness, E. y Sønju, T. (1995). Analysis of dependent survival data applied to lifetimes of amalgam fillings. *Statistics in Medicine* **14**, 1819-1829.
- [2] Allison, P.D. (1995). *Survival analysis using SAS System: A practical guide*. Cary, N.C.: SAS Institute, Inc.

- [3] Andersen, P.K., Borgan, Ø., Gill, R.D. y Keiding, N. (1993). *Statistical models based on counting processes*. N.Y.: Springer-Verlag.
- [4] Armitage, P. y Colton, T. (editores) (1998). *Encyclopedia of Biostatistics*. N.Y.: John Wiley and Sons.
- [5] Borges, R.E. (2002). *Análisis de Supervivencia Aplicado a un Caso de Diálisis Renal: Diálisis Peritoneal en el Hospital Clínico Universitario de Caracas y Hemodiálisis en el Hospital de Clínicas Caracas, 1980-2000*. Tesis de M.Sc. en Estadística Aplicada, Mérida, Venezuela: Instituto de Estadística Aplicada y Computación, Universidad de Los Andes. (Disponible en: <http://tesis.saber.ula.ve>)
- [6] Cox, D.R. y Oakes, D. (1972). *Analysis of Survival Data*. Londres: Chapman & Hall.
- [7] Crowder, M. (2001). *Classical Competing Risks*. Boca Ratón, Florida: Chapman & Hall.
- [8] Crowley, J. y Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72**, 27-35.
- [9] Dabrowska, D. (1998). Multivariate Survival Analysis. En: Armitage, P. y Colton, T. (editores). *Encyclopedia of Biostatistics*. N.Y.: John Wiley and Sons. pp. 2927-2933.
- [10] Danahy, D.T. et al (1977). Sustained hemodynamic and antianginal effect of high dose oral isosorbide dinitrate. *Circulation* **55**, 381-387.
- [11] Fleming, T.R. (1992). Evaluating therapeutic interventions: Some issues and experiences (with discussion). *Statistical Science* **7**, 428-456.
- [12] Freireich, E.J. et al. (1963). The effect of 6-Mercaptopurine on the duration of steroid-induced remissions in acute leukemia: model for evaluation of other potentially useful therapy. *Blood* **21**, 699-716.
- [13] Gail, M.H., Santner, T.J. y Brown, C.C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36**, 255-266.
- [14] Gall, M.A., Borch-Johnsen, K., Hougaard, P., Nielsen, P.V. y Parving, V.V. (1995). Albuminuria and poor glycemic control predict mortality in NIDDM. *Diabetes* **44**, 1303-1309.
- [15] Gall M.A., Hougaard, P., Borch-Johnsen, K., y Parving, V.V. (1997). Risk factor for development of incipient and overt diabetic nephropathy inpatients with non-insulin dependent diabetes mellitus: prospective, observational study. *British Medical Journal* **314**, 783-788.

- [16] Hauge, M., Harvald, B., Fischer, M., Gotlieb-Jensen, K., Juel-Nielsen, N., Reabild, I., Shapiro, R. y Videbeck, T. (1968). The Danish twin register. *Acta Geneticae Medicae et Gemellologiae* **17**, 315-331.
- [17] Hoel, D.G. y Walburg, H.E. (1972). Statistical analysis of survival experiments. *Journal of the National Cancer Institute* **49**, 361-372.
- [18] Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-396.
- [19] Hougaard, P. (1987). Modelling multivariate survival. *Scandinavian Journal of Statistics* **14**, 291-304.
- [20] Hougaard, P. (1998). Frailty. En: Armitage, P. y Colton, T. (editores). *Encyclopedia of Biostatistics*. N.Y.: John Wiley and Sons. pp. 1572-1577.
- [21] Hougaard (2000). *Analysis of Multivariate Survival Data*. N.Y.: Springer-Verlag.
- [22] Hougaard, P., Harvald, B. y Holm, B.V. (1992a). Measuring the similarities between the lifetime of adult Danish twins born between 1881-1930. *Journal of the American Statistical Association* **87**, 1881-1930.
- [23] Hougaard, P., Harvald, B. y Holm, B.V. (1992b). Assessment of dependence of the life times of twins. En: Klein, J.P. y Goel, P.K. (editores). *Survival Análisis: State of the Art*. Dordrecht, Holanda: Kluwer Academic Publisher, pp. 77-97.
- [24] Huster, W.J., Brookmeyer, R. y Self, S.G. (1989). Modelling paired survival data with covariates. *Biometrics* **45**, 145-156.
- [25] Kalbfleisch, J.D. y Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. N.Y.: John Wiley & Sons, Inc.
- [26] Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795-806.
- [27] Klein, J.P. y Goel, P.K. (editores) (1992). *Survival Análisis: State of the Art*. Dordrecht, Holanda: Kluwer Academic Publisher.
- [28] Klein, J.P. y Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.
- [29] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. N.Y.: John Wiley & Sons, Inc.
- [30] Liang, K.Y., Self, S.G. y Chang, Y.C. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society, Series B.* **55**, 441-453.

- [31] Lin, D.Y. y Fleming, T.R. (editores) (1997). *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*. N.Y.: Springer-Verlag (Lecture Notes in Statistics 123).
- [32] Mantel, N., Bohidar, N.R. y Ciminera, J.L. (1977). Mantel-Haenszel analysis of litter-matched time-to-response data with modifications for recovery of interlitter information. *Cancer Research* **37**, 3863-3866.
- [33] McGilchrist, C.A. y Aisbeth, C.W. (1991). Regression with frailty in survival analysis. *Biometrics* **47**, 461-466.
- [34] Moertel, C.G., Fleming, T.R., McDonald, J.S., Haller, D.G., Laurie, J.A., Goodman, P.J., Ungerleider, J.S., Emerson, W.A., Tormey, D.C., Glick, J.H., Veeder, M.H. y Mailliard, J.A. (1990). Levamisole and Fluorouracil or adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine* **322**, 352-358.
- [35] Nielsen, G.G., Gill, R.D., Andersen, P.K. y Sørensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25-43.
- [36] Pickles, A. Chrouchley, R., Simonoff, E., Eaves, L., Meyer, J. Rutter, M., Hewitt, J. y Silberg, J. (1994). Survival models for development genetic data: age of onset of puberty and antisocial behaviour in twins. *Genetic Epidemiology* **11**, 155-170
- [37] Sørensen, T.I.A., Nielsen, G.G., Andersen, P.K. y Teasdale, T.W. (1988). Genetic and environmental influences on premature death in adult adoptees. *New England Journal of Medicine* **318**, 727-732.
- [38] Thall, P.F. y Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657-671.
- [39] Therneau, T.M. y Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. N.Y.: Springer-Verlag.
- [40] Vaupel, J.W., Manton, K.G. y Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439-454.