

# Análisis de supervivencia de pacientes con diálisis peritoneal

Survival Data Analysis of Patients with Peritoneal Dialysis

RAFAEL EDUARDO BORGES P.\*

---

## Resumen

Se realiza un análisis de supervivencia de pacientes que acudían al Servicio de Diálisis Peritoneal del Hospital Clínico Universitario de Caracas, Venezuela, entre los años 1980 y 1997, utilizando la muerte como evento de interés. El análisis incluye: estimaciones de las funciones de supervivencia mediante el estimador de Kaplan–Meier, obtención del mejor modelo semi-paramétrico de riesgos proporcionales (modelo de Cox), verificación de los supuestos y análisis de residuos. El modelo de Cox incluye diabetes, edad y el índice de Quetlet como covariables.

**Palabras Claves:** Análisis de supervivencia, estimador de Kaplan y Meier, modelo de riesgos proporcionales (modelo de Cox), diálisis peritoneal.

## Abstract

A survival analysis in patients that assisted to the peritoneal dialysis service of the *Hospital Clínico Universitario de Caracas, Venezuela*, between 1980 and 1997, using death as event of interest is presented. Estimation of the survival function by the Kaplan-Meier estimator, Semi-parametric proportional hazard models (Cox models), verification of the assumptions of the models and residual analysis are included. In the Cox model, diabetes, age and Quetlet's index are used as covariates.

**Keywords:** Survival Analysis, Kaplan-Meier estimator, proportional hazard model (Cox model), continuous ambulatory peritoneal dialysis.

## 1. Introducción

Las aplicaciones del análisis de supervivencia en diálisis renal no son nuevas. Sin embargo, la mayoría de ellas se limitan a la estimación de funciones

---

\*Profesor Agregado, Escuela de Estadística, Facultad de Ciencias Económicas y Sociales, Universidad de Los Andes, Mérida 5101, Venezuela. E-mail: borgesr@ula.ve

de supervivencia o de riesgo y al ajuste de modelos de regresión para encontrar predictores de mortalidad o razones de riesgo, sin la verificación de los supuestos de los modelos (Hutchinson, Thomas, Lemieux & Harvey 1984, Burton & Walls 1987, Maiorca, Vonesh, Cancarini, Cantaluppi, Manili, Brunori, Camerini, Sèller & Strada 1988, Maiorca, Vonesh, Cavalli, De Vecchi, Giangrande, La Greca, Scarpione, Bragantini, Cancarini, Cantaluppi, Castelnovo, A., Poisetti & Viglino 1991, Held, Port, Turrenne, Gaylin, Hamburger & Wolfe 1994, Maggiore, Nigrelli, Cicarelli, Grimaldi, Rossi & Michelassi 1996, Iseki, Miyasato, Tokuyama, Nishime, Ueshara, Shiohira, Sunagawa, Yoshihara, Yoshi, Toma, Kowatari, Wake, Oura & Fukiyama 1997, Woods, Port, Orzoul, Buoncristiani, Wolfe & Held 1998).

El enfoque basado en procesos de conteo ha ampliado las posibilidades de los modelos de supervivencia (Fleming & Harrington 1991, Andersen, Borgan, Gill & Keiding 1993). Therneau & Grambsch (2000) introducen una metodología basada en residuos que permite verificar los supuestos de los modelos de regresión, incluida en el software comercial reciente. El objetivo de este trabajo es presentar un análisis de supervivencia clásico y verificar los supuestos de los modelos de regresión basados en las técnicas más nuevas disponibles (Borges 2002).

## 2. Datos

Los datos analizados corresponden a 246 pacientes en diálisis peritoneal (DPA) que acudían al Servicio del Hospital Clínico Universitario de Caracas entre 1980 y 1997. Se hizo un seguimiento a los pacientes desde el comienzo de sus sesiones de diálisis hasta alcanzar la muerte como evento de interés, o hasta la terminación del estudio, por lo que algunas observaciones resultan censuradas. En el análisis inicial se incluyeron 100 covariables dicotómicas y 16 continuas.

## 3. Análisis estadístico

Se estimó la función de supervivencia mediante el estimador de Kaplan & Meier (1958). Se ajustaron varios modelos de Cox (1972) para obtener las covariables significativas, eliminando las variables no significativas mediante el procedimiento paso a paso hacia atrás. Se aplicó un análisis de residuos a los modelos definitivos para verificar los supuestos del modelo de Cox, según lo propuesto por Therneau & Grambsch (2000). Para los cálculos se utilizó el paquete S-PLUS.

## 4. Elementos de la teoría de análisis de supervivencia

El análisis de supervivencia tiene como objeto de estudio el tiempo de seguimiento hasta la ocurrencia de un evento de interés y cobra vital importancia cuando existen observaciones censuradas. Existen varios tipos de censura: por la

derecha, por la izquierda y por intervalos. Para tener un panorama general de los distintos tipos de censura véase Andersen et al. (1993). En este caso, trataremos con la censura más común, la censura por la derecha: se presenta cuando hasta el último instante en que se ha seguido al individuo todavía no ha ocurrido el evento de interés.

## 4.1. Definiciones básicas

### 4.1.1. Función de supervivencia

La función de supervivencia se define como la probabilidad de que una persona sobreviva (no le ocurra el evento de interés) al menos hasta el tiempo  $t$ . Una definición más formal puede darse de la siguiente manera: 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  $S(t)$  es:

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

### 4.1.2. Función de razón de riesgos (hazard rate)

La función de razón de riesgos o tasa instantánea de fallas  $\lambda(t)$  se define como el cociente entre la función de densidad y la función de supervivencia:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

Se interpreta como la probabilidad de que a un individuo le ocurra el evento de interés en la siguiente unidad de tiempo  $\Delta t$  dado que ha sobrevivido hasta el tiempo  $t$ .

Dicha función proviene de la tasa media de fallas: dada la probabilidad condicional de fallas en el período  $(t; t + \Delta t)$ , dado que la persona sobrevive en el período  $(0; t)$ , la tasa media de fallas (TMF) se define como:

$$\text{TMF} = \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)}$$

Tomando límites para  $\Delta t \rightarrow 0$ , queda:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \text{TMF} = \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)}$$

La función de riesgo acumulada  $\Lambda(t)$  se define como:

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log S(t)$$

Los datos de supervivencia suelen presentarse en la forma  $(t_i, \delta_i)$  donde  $t_i$  es el tiempo de observación y,  $\delta_i = 0$  si la observación es censurada y  $\delta_i = 1$  cuando se observa la ocurrencia del evento de interés.

## 4.2. Estimador de Kaplan y Meier

El estimador de la función de supervivencia más utilizado es el de Kaplan & Meier (1958):

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \frac{r(t_i) - d(t_i)}{r(t_i)}$$

donde  $r(t_i)$  y  $d(t_i)$  son el número de individuos en riesgo y el número de muertes (o de ocurrencia del evento de interés en el momento  $t_i$ ).

La varianza del estimador de Kaplan y Meier se obtiene a través de la fórmula de Greenwood (1926):

$$V(\hat{S}_{KM}(t)) = \hat{S}_{KM}^2(t) \sum_{t_i \leq t} \frac{d(t_i)}{r(t_i)[r(t_i) - d(t_i)]}$$

El intervalo de confianza calculado por defecto por los programas estadísticos es el de identidad o de escala plana, dado, para un nivel de confianza del 90 %, por:

$$\hat{S}_{KM}^2(t) \pm 1.645 \text{ ee}(\hat{S}_{KM}(t))$$

donde  $\text{ee}(\hat{S}_{KM}(t))$  es el error estándar de estimación del estimador de Kaplan y Meier.

## 4.3. El modelo de regresión de Cox

El modelo de regresión de Cox (1972) es el más utilizado para datos de supervivencia en el área médica. En este modelo 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  $t$ . Se dice que es un modelo semiparamétrico debido a que incluye una parte paramétrica y otra no paramétrica:

- i) La parte paramétrica es  $r_i(t) = e^{\beta' Z_i(t)}$ , llamada puntaje de riesgo (*risk score*) y  $\beta$  es el vector de parámetros de la regresión.
- ii) La parte no paramétrica es  $\lambda_0(t)$ , llamada función de riesgo base y es una función arbitraria no especificada.

El modelo de regresión de Cox es también llamado modelo de riesgos proporcionales debido a que el cociente entre el riesgo para dos sujetos con el mismo vector de covariables es constante sobre el tiempo, es decir:

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

Si ha ocurrido una muerte en el tiempo  $t^*$ , entonces la verosimilitud de que la muerte le ocurra al  $i$ -ésimo individuo y no a otro es:

$$L_i(\beta) = \frac{\lambda_0(t^*) r_i(t^*)}{\sum_j Y_j(t^*) \lambda_0(t^*) r_j(t^*)} = \frac{r_i(t^*)}{\sum_j Y_j(t^*) r_j(t^*)}$$

El producto  $L(\beta) = \prod L_i(\beta)$  se llama la verosimilitud parcial y fue introducida por Cox (1972). La maximización de  $\log(L(\beta))$  da una estimación para  $\beta$  sin necesidad de estimar el parámetro de ruido  $\lambda_0(t)$ .

#### 4.4. Contrastes de hipótesis para el modelo de Cox

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.

##### 4.4.1. Test de razón de verosimilitud

El primero de los contrastes es el denominado test de razón de verosimilitud y es el que presenta una mayor confiabilidad. Este test se define como:

$$2\left\{\log(L(\beta_0)) - \log(L(\hat{\beta}))\right\}$$

donde  $\beta_0$  son los valores iniciales de los coeficientes y  $\hat{\beta}$  es la solución luego de ajustar el modelo. Este es el test que presentan por defecto los paquetes computacionales.

##### 4.4.2. Test de Wald

El segundo de los contrastes es conocido como el test de Wald y es quizás el más natural debido a que proporciona un contraste por variables en vez de una medida de significación global. El estadístico de contraste se define mediante:

$$(\hat{\beta} - \beta_0)' \hat{\Sigma}_{\hat{\beta}}^{-1} (\hat{\beta} - \beta_0)$$

donde  $\hat{\Sigma}_{\hat{\beta}}$  es la matriz de varianzas y covarianzas estimada.

##### 4.4.3. Test de los puntajes (score test)

El tercer contraste es el conocido como test de los puntajes (*score test*), definido como  $U'IU$ , donde  $U$  es el vector de derivadas del  $\log(L(\beta))$  dado por:

$$U(\beta) = \sum_{i=1}^n \int_0^{\infty} [Z_i(t) - \bar{Z}(\beta, t)] dN_i(t)$$

$I$  es la matriz de información:

$$I(\beta) = \sum_{i=1}^n \int_0^{\infty} \frac{\sum_j Y_j(t) r_j(t) [Z_i(t) - \bar{Z}(\beta, t)] [Z_i(t) - \bar{Z}(\beta, t)]'}{\sum_j Y_j(t) r_j(t)} dN_i(t)$$

y  $\bar{Z}(\beta, t)$  es la media de las covariables para aquellos individuos todavía en riesgo en el tiempo  $t$ :

$$\bar{Z}(\beta, t) = \frac{\sum_j Y_j(t) r_j(t) Z_i(t)}{\sum_i Y_i(t) r_i(t)}$$

## 4.5. Interpretación del modelo de Cox

La interpretación del modelo de Cox no se hace directamente a través de su coeficiente estimado sino de su exponencial,  $\exp(\hat{\beta})$ .

### 4.5.1. Interpretación para covariables dicotómicas

Para cada covariable dicotómica,  $\exp(\hat{\beta})$  es un estimador de la razón de riesgos (*hazard ratio*) y se interpreta como la cantidad de riesgo que se tiene con la presencia de la covariable en relación a la ausencia de la covariable. Los intervalos de confianza del 90 % para  $\exp(\hat{\beta})$  se obtienen mediante:

$$\exp(\hat{\beta} \pm 1.645 ee(\hat{\beta}))$$

donde  $ee(\hat{\beta})$  es el error estándar de  $\hat{\beta}$ .

### 4.5.2. Interpretación para covariables continuas

Para el caso de covariables continuas,  $\exp(\hat{\beta})$  representa la razón de riesgos (*hazard ratio*) al incrementar en una unidad la covariable.

En el caso de las covariables continuas suele resultar más interesante estimar la razón de riesgos al incrementar la covariable en  $c$  unidades y esto se hace mediante  $\exp(c\hat{\beta})$ , siendo su intervalo de confianza del 90 % de la forma:

$$\exp(c\hat{\beta} \pm 1.645 |c| ee(\hat{\beta}))$$

Para una explicación más detallada puede verse Hosmer & Lemeshow (1999).

## 4.6. Estudio de residuos en el análisis de supervivencia

Una de las ventajas que han surgido del enfoque del análisis de supervivencia es la posibilidad de efectuar análisis de residuos (Andersen et al. 1993, Fleming & Harrington 1991, Therneau & Grambsch 2000, Therneau, Grambsch & Fleming 1990).

Los residuos se pueden utilizar para:

1. descubrir la forma funcional correcta de un predictor continuo.
2. identificar los sujetos que están pobremente pronosticados por el modelo.
3. identificar los puntos o individuos de influencia.
4. verificar el supuesto de riesgo proporcional.

Existen cuatro tipos de residuos de interés en el modelo de Cox: los de martingala, los de desvíos (*deviances*), los de puntaje (*score*) y los de Schoenfeld. De ellos pueden derivarse otros dos: los *dfbetas* y los residuos escalados de Schoenfeld. A continuación explicaremos brevemente cada uno de ellos.

#### 4.6.1. Residuos de martingala

Los residuos de martingala se definen como:

$$\hat{M}_i(t) = N_i(t) - \hat{E}_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\beta' Z_i(s)} d\hat{\Lambda}_0(\beta, s)$$

donde  $\hat{\Lambda}_0(\beta, s)$  es el estimador del riesgo base de Breslow (o de Tsiatis o de Nelson y Aalen) definido como:

$$\hat{\Lambda}_0(\beta, s) = \int_0^s \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) e^{\beta' Z_i(s)}}$$

y están basados en la martingala de un proceso de conteo para el  $i$ -ésimo individuo,  $M_i(t) = N_i(t) - E_i(t)$ , definida mediante:

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\beta' Z_i(s)} \lambda_0(s) ds$$

Los residuos de martingala son muy asimétricos y con una cola muy larga hacia la derecha, particularmente para datos de supervivencia para un solo evento. Se usan para estudiar la forma funcional de una covariable continua.

#### 4.6.2. Residuos de desvíos (*deviances*)

Los residuos de desvíos se obtienen mediante una transformación de normalización de los de martingala y son similares en forma a los residuos de desvíos (*deviances*) en la regresión de Poisson. Si todas las covariables son fijas en el tiempo, los residuos de desvíos toman la forma:

$$d_i = \text{signo}(\hat{M}_i) * \sqrt{-\hat{M}_i - N_i \log((N_i - \hat{M}_i)/N_i)}$$

Una expansión de Taylor de un término muestra que

$$d_i \approx \frac{N_i - \hat{E}_i}{\sqrt{\hat{E}_i}}$$

que es formalmente equivalente a los residuos de Pearson de los modelos lineales generalizados. Los residuos de desvíos se utilizan para la detección de valores atípicos (*outliers*).

#### 4.6.3. Residuos de puntajes (*scores*)

Los residuos de puntajes se definen como:

$$U_{ij} = U_{ij}(\hat{\beta}, \infty)$$

donde  $U_{ij}(\beta, t)$ ,  $j = 1, \dots, p$  son las componentes del vector fila de longitud  $p$  obtenido a través del proceso de puntaje para el  $i$ -ésimo individuo:

$$U_i(\beta) = \int_0^t [Z_i(t) - \bar{Z}(\beta, t)] dN_i(t)$$

Los residuos de puntajes se utilizan para verificar la influencia individual y para la estimación robusta de la varianza.

#### 4.6.4. Residuos de Schoenfeld

Los residuos de Schoenfeld se definen como la matriz

$$s_{ij}(\beta) = Z_{ij}(t_i) - \bar{Z}_j(\beta, t_i)$$

que tiene una fila por muerte y una columna por covariable, donde  $i$  y  $t_i$  son los individuos y el tiempo de ocurrencia del evento respectivamente. Estos residuos son útiles para la verificación del supuesto de riesgo proporcional en el modelo de Cox.

## 5. Resultados

### 5.1. Estimador de Kaplan y Meier

Los valores obtenidos con el estimador de Kaplan y Meier con el intervalo, el error estándar y el intervalo de confianza del 90% puede verse en la tabla 1.



Tabla 1: Modelos de Cox para DPA

| tiempo | n.riesgo | n.eventos | sobrevida | err.est. | LCI 90 % | LCS 90 % |
|--------|----------|-----------|-----------|----------|----------|----------|
| 0      | 246      | 2         | 0.992     | 0.00573  | 0.982    | 1.000    |
| 1      | 240      | 1         | 0.988     | 0.00704  | 0.976    | 0.999    |
| 3      | 228      | 4         | 0.970     | 0.01102  | 0.952    | 0.989    |
| 4      | 221      | 1         | 0.966     | 0.01182  | 0.947    | 0.986    |
| 5      | 215      | 1         | 0.962     | 0.01259  | 0.941    | 0.982    |
| 6      | 209      | 1         | 0.957     | 0.01334  | 0.935    | 0.979    |
| 7      | 202      | 1         | 0.952     | 0.01409  | 0.929    | 0.976    |
| 8      | 197      | 1         | 0.947     | 0.01483  | 0.923    | 0.972    |
| 9      | 193      | 1         | 0.942     | 0.01554  | 0.917    | 0.968    |
| 10     | 188      | 1         | 0.937     | 0.01625  | 0.911    | 0.965    |
| 11     | 180      | 3         | 0.922     | 0.01831  | 0.892    | 0.952    |
| 12     | 171      | 1         | 0.916     | 0.01898  | 0.886    | 0.948    |
| 13     | 161      | 1         | 0.911     | 0.01970  | 0.879    | 0.944    |
| 14     | 151      | 4         | 0.887     | 0.02257  | 0.850    | 0.925    |
| 15     | 144      | 1         | 0.880     | 0.02324  | 0.843    | 0.920    |
| 17     | 135      | 3         | 0.861     | 0.02532  | 0.820    | 0.904    |
| 19     | 124      | 1         | 0.854     | 0.02605  | 0.812    | 0.898    |
| 20     | 119      | 2         | 0.840     | 0.02752  | 0.796    | 0.886    |
| 21     | 115      | 3         | 0.818     | 0.02956  | 0.770    | 0.868    |
| 22     | 110      | 3         | 0.795     | 0.03143  | 0.745    | 0.849    |
| 23     | 104      | 1         | 0.788     | 0.03205  | 0.737    | 0.842    |
| 25     | 94       | 1         | 0.779     | 0.03278  | 0.727    | 0.835    |
| 26     | 90       | 1         | 0.771     | 0.03354  | 0.717    | 0.828    |
| 28     | 81       | 1         | 0.761     | 0.0345   | 0.7066   | 0.820    |
| 30     | 78       | 2         | 0.742     | 0.0362   | 0.6844   | 0.804    |
| 31     | 75       | 4         | 0.702     | 0.0393   | 0.6403   | 0.770    |
| 33     | 63       | 1         | 0.691     | 0.0403   | 0.6278   | 0.760    |
| 34     | 59       | 1         | 0.679     | 0.0412   | 0.6147   | 0.751    |
| 37     | 53       | 2         | 0.654     | 0.0435   | 0.5859   | 0.729    |
| 39     | 50       | 1         | 0.641     | 0.0445   | 0.5713   | 0.718    |
| 42     | 43       | 1         | 0.626     | 0.0459   | 0.5545   | 0.706    |
| 47     | 38       | 1         | 0.609     | 0.0476   | 0.5358   | 0.693    |
| 52     | 33       | 1         | 0.591     | 0.0496   | 0.5145   | 0.678    |
| 55     | 31       | 1         | 0.572     | 0.0515   | 0.4929   | 0.663    |
| 59     | 26       | 2         | 0.528     | 0.0562   | 0.4430   | 0.629    |
| 60     | 22       | 1         | 0.504     | 0.0585   | 0.4161   | 0.610    |
| 61     | 21       | 1         | 0.480     | 0.0604   | 0.3899   | 0.590    |
| 65     | 18       | 1         | 0.453     | 0.0627   | 0.3609   | 0.569    |
| 77     | 11       | 1         | 0.412     | 0.0692   | 0.3124   | 0.543    |
| 96     | 7        | 1         | 0.353     | 0.0805   | 0.2426   | 0.514    |
| 110    | 4        | 2         | 0.177     | 0.0970   | 0.0715   | 0.436    |

La tabla 1 incluye la siguiente información:

|                  |  |
|------------------|--|
| <i>tiempo</i>    | número de meses de seguimiento.  |
| <i>n.riesgo</i>  | número de individuos en riesgo antes del tiempo.                                   |
| <i>n.eventos</i> | número de muertes entre en tiempo y el siguiente mes en donde ocurre una muerte.   |
| <i>sobrevida</i> | probabilidad de que un individuo sobreviva por un número de meses mayor al tiempo. |
| <i>err.est.</i>  | error estándar de la supervivida.  |
| <i>LCI 90 %</i>  | límite de confianza inferior del 90 % para la supervivida.                         |
| <i>LCS 90 %</i>  | límite de confianza superior del 90 % para la supervivida.                         |

Puede verse que el 75 % de los pacientes en diálisis peritoneal sobrevive hasta los 30 meses en diálisis, el 50 % logra sobrevivir hasta 61 meses, el 25 % sobrevive hasta 110 meses y el 17.7 % de los pacientes sobrevive más de 110 meses. El tiempo se supervivida medio se ubicó en 67.2 meses, con un error estándar de 4.46 meses.

## 5.2. Modelo de Cox

El mejor modelo de Cox ajustado para DPA y muerte como evento de interés se muestra en la tabla 2. Puede afirmarse que las variables edad, índice de Quetelet y diabetes son significativas al 10 %, debido a que los  $p$ -valores obtenidos son todos menores que 0.10.

Tabla 2: Modelo de Cox para DPA y muerte como evento de interés.

| Covariables        | Coefficiente | $p$ -valor |
|--------------------|--------------|------------|
| Edad               | 0.0315       | 0.0011     |
| Índice de Quetelet | -0.0969      | 0.013      |
| Diabetes           | 0.5492       | 0.087      |

Este modelo resulta significativo por cualquiera de los tres criterios para un 10 % de significación, debido que los  $p$ -valores son todos menores que 0.10. Para el test de razón de verosimilitud se obtuvo un  $p$ -valor de 0.000308, para el test de Wald fue de 0.000229 y para el test de puntajes fue de 0.000184.

La figura 1 muestra el estimador de Kaplan y Meier de la función de supervivencia y el ajuste mediante el modelo de Cox. En esta figura puede observarse que el ajuste obtenido mediante el modelo de Cox se ubica sistemáticamente por encima de la estimación de Kaplan y Meier.

## 5.3. Interpretación de los coeficientes estimados

La interpretación de los coeficientes debe hacerse a través de la evaluación de la función exponencial evaluada sobre los coeficientes estimados para cada covariable y esta interpretación difiere para covariables continuas y discretas.

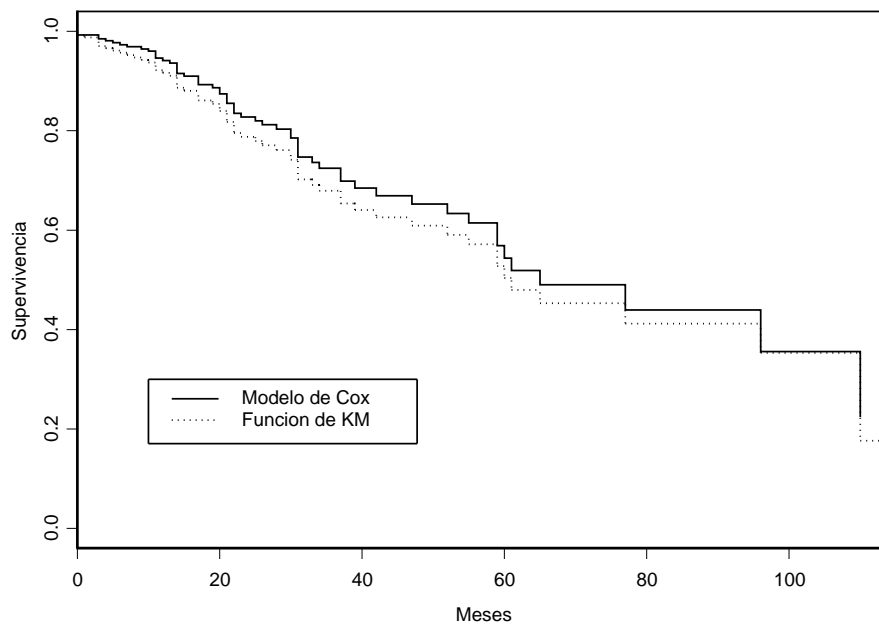


Figura 1: Comparación del ajuste del modelo de Cox y el estimador de KM para DPA según meses.

### 5.3.1. Interpretación para la edad

El exponencial de coeficiente estimado para la edad es 1.032. Así, por cada año que aumenta la edad del paciente, el riesgo de morir por causas asociadas a la diálisis peritoneal es de 1.032 veces la de edad menor.

Por ser una variable continua, la interpretación para la edad puede hacerse para un período de distinto tamaño, utilizando a  $\exp(c\hat{\beta})$  en vez de  $\exp(\hat{\beta})$ , donde  $c$  es el número de años contenidos en el período.

Por ejemplo, si tomamos períodos de 5 años de edad, se obtendría un riesgo de 1.171, lo que significa que al aumentar la edad del paciente en cinco años, el riesgo de morir por causas asociadas a la diálisis peritoneal es 1.171 veces la de la edad menor.

Análogamente, si tomamos períodos de 10 años de edad, se obtendría un riesgo de 1.370, lo que significa que al aumentar la edad del paciente en diez años, el riesgo de morir por causas asociadas a la diálisis peritoneal es 1.370 veces la de la edad menor.

### 5.3.2. Interpretación para el índice de Quetelet

La interpretación del índice de Quetelet es análoga a la de la edad, el exponencial del coeficiente estimado en este caso fue de 0.908, valor que se interpreta

de la siguiente manera: al aumentar el índice de Quetelet es una unidad, el riesgo es 0.908 veces en comparación con la unidad menor. Este resultado indica que el índice de Quetelet pareciera ser un factor de protección, en lugar de un factor de riesgo, es decir, que mientras mayor es el índice, menor es la mortalidad por causas asociadas a la diálisis peritoneal.

### 5.3.3. Interpretación para la diabetes

Para el caso de la diabetes, la interpretación es distinta: en este caso el exponencial del coeficiente estimado es 1.732, que se interpreta así: la presencia de diabetes aumenta el riesgo de muerte por causas asociadas a diálisis peritoneal en 1.732 veces, es decir, que un individuo con diabetes tiene 1.732 veces más riesgo de morir por causas asociadas a la diálisis peritoneal que un individuo que no tenga diabetes.

## 5.4. Análisis de residuos

En esta sección se presenta el análisis de residuos mediante dispositivos gráficos incluidos en la figura 2 donde:

- (a) Es el gráfico para el test de riesgos proporcionales para diabetes.
- (b) Es el gráfico para el test de riesgos proporcionales para edad.
- (c) Es el gráfico para el test de riesgos proporcionales para índice de Quetelet.
- (d) Es el gráfico de influencias para diabetes.
- (e) Es el gráfico de influencias para edad.
- (f) Es el gráfico de influencias para índice de Quetelet.
- (g) Es el gráfico de los devíos (deviances) del modelo.
- (h) Es el gráfico para la verificación de la adecuación de la forma funcional de la edad.
- (i) Es el gráfico para la verificación de la adecuación de la forma funcional del índice de Quetelet.

### 5.4.1. Supuestos de riesgos proporcionales

La verificación de los supuestos de riesgos proporcionales puede verse mediante las partes (a), (b) y (c) de la figura 2. En estos gráficos no se observa una violación del supuesto en cada una de las covariables. Sin embargo, para el caso de la edad, se observa quizás un patrón cíclico muy atenuado, lo que da indicios de la necesidad de ajustar un modelo de Cox con covariables dependientes del tiempo, pudiendo

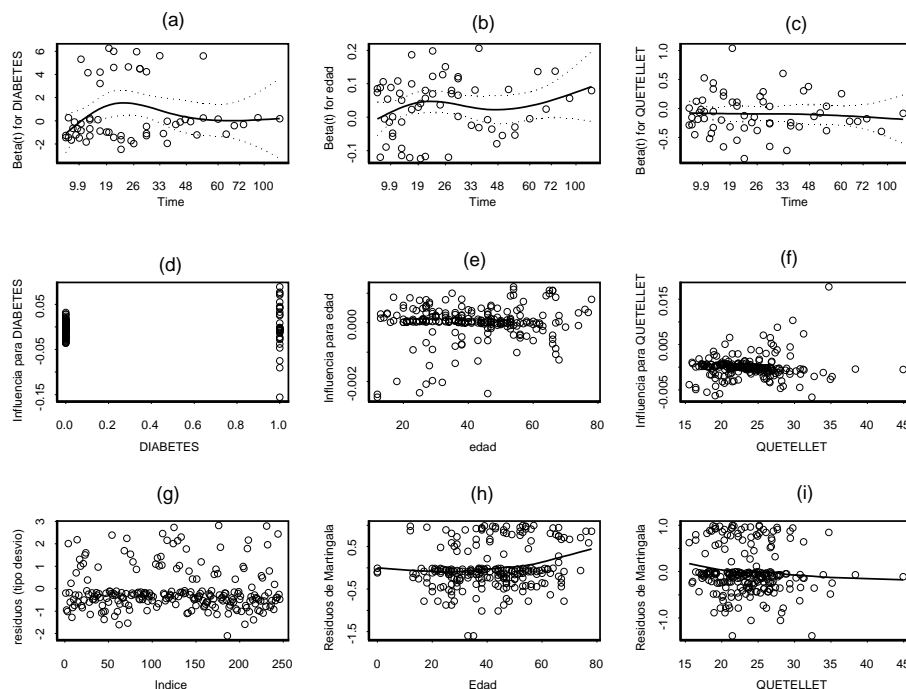


Figura 2: Verificación de los supuestos del modelo de Cox.

en este caso utilizarse los modelos frágiles (frailty models), debido a la variación temporal de los residuos de Schoenfeld.

Los modelos frágiles son modelos de efectos aleatorios para datos de supervivencia introducidos por Vaupel, Manton & Stallard (1979) y han sido utilizados en los modelos de riesgos proporcionales con covariables dependientes del tiempo. Esta no es la única aplicación de los modelos frágiles, una revisión exhaustiva puede verse en Hougaard (1995) o en Liang, Self, Bandeen-Roche & Zeger (1995).

La verificación del supuesto de riesgos proporcionales puede efectuarse a través de un contraste de hipótesis, donde la hipótesis nula está asociada al cumplimiento del supuesto de riesgos proporcionales. Los resultados de este contraste indican que no se viola el supuesto de riesgos proporcionales para ninguna de las tres covariables. Los  $p$ -valores asociados a este contraste para diabetes, edad e índice de Quetlet son 0.776, 0.305 y 0.633, respectivamente, observándose que todos son mayores que 0.10, con lo que no se estaría rechazando la hipótesis de riesgos proporcionales para ninguna de las covariables.

Este contraste permite verificar la violación global del supuesto de riesgos proporcionales de todas las covariables. En este caso se obtiene un  $p$ -valor de 0.71, y por ser éste mayor que 0.10, no se rechazaría la hipótesis nula de cumplimiento conjunto del riesgo proporcional de las tres covariables.

### 5.4.2. Influencia de individuos en la estimación de los coeficientes

La contribución de los individuos en la estimación de los coeficientes puede verse a través de las partes (d), (e) y (f) de la figura 2. En estos gráficos se puede observar que para diabetes y edad, no existen individuos que estén influyendo en la estimación de sus respectivos coeficientes. Para el índice de Quetelet, probablemente existe un individuo ubicado en la parte superior que está influyendo en la estimación de su coeficiente; se trata del individuo # 6.

### 5.4.3. Influencia de individuos en la estimación del modelo

En la parte (g) de la figura 2 no se observa ningún individuo que esté influyendo en la estimación del modelo, ya que el patrón es el de una nube de puntos.

### 5.4.4. Forma funcional de las covariables continuas

En las partes (h) e (i) de la figura 2 se observa que la forma funcional es correcta en el modelo que puede verse al tener las funciones suavizadas la apariencia de líneas rectas, tanto para la edad, como para el índice de Quetelet. Eso quiere decir que no hace falta efectuar transformaciones para estas covariables.

## 6. Conclusiones

El análisis de supervivencia clásico es adecuado para la estimación de funciones de supervivencia y el ajuste de modelos de regresión para la obtención de covariables significativas, lo cual queda evidenciado en este trabajo.

La incorporación del enfoque de procesos de conteo al análisis de supervivencia ha permitido el desarrollo de nuevas herramientas. En este caso se ha considerado sólo el análisis de los residuos para la verificación de los supuestos del modelo de Cox. Los aportes de este enfoque al análisis de supervivencia son muchos.

Puede concluirse que para el caso de los pacientes que acudían al Servicio de diálisis peritoneal del Hospital Clínico Universitario de Caracas, Venezuela entre los años 1980 y 1997, las covariables significativas en el modelo de Cox fueron la diabetes, la edad y el índice de Quetelet. Estas covariables son las que estarían modificando el riesgo de muerte en los pacientes en diálisis peritoneal.

Se concluye además que el modelo de riesgos proporcionales presentado es adecuado ya que todos los supuestos se verifican.

## 7. Discusión

A pesar de que el análisis de supervivencia es una herramienta muy poderosa para modelar datos de evento-tiempo y la más adecuada para datos censurados, se encuentra poco vinculada a los programas de estudios en estadística en nuestros

países y no hay duda en que debe incluirse de manera formal en los programas de estudios de pregrado y postgrado.

Debemos resaltar que el campo de acción del análisis de supervivencia no está sólo vinculada con el área médica, sino con cualquier área en donde se quiera determinar las funciones del tiempo transcurrido desde un instante de comienzo del seguimiento de un conjunto de individuos hasta la ocurrencia de un evento de interés, y en caso de no observar el evento de interés se tienen las observaciones censuradas.

En este trabajo sólo se presentan la estimación de las funciones de supervivencia, la obtención de covariables predictoras de la función de supervivencia y la verificación de los supuestos del modelo de Cox. Un paso más adelante a lo alcanzado hasta este punto lo constituye la estimación de las razones de riesgo para cada una de las modalidades de las covariables dicotómicas. Otra opción para continuar con el presente trabajo es la obtención de modelos que contemplen variación temporal del supuesto de riesgo proporcional, pudiéndose utilizar en este caso la metodología de los modelos frágiles.

El análisis presentado fue realizado utilizando en software estadístico *S-PLUS* pero éste puede llevarse a cabo mediante otras herramientas como *R* (R Development Core Team 2005).

**Agradecimientos** Agradezco al Prof. Giampaolo Orlandoni, tutor, y al Dr. Pablo Amair, jurado de mi tesis de maestría en Estadística Aplicada de donde se tomó una parte como base para este artículo.

*Recibido: 30 de Septiembre de 2004*

*Aceptado: 6 de Julio de 2005*

## Referencias

- Andersen, P., Borgan, O., Gill, R. & Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag, N.Y.
- Borges, R. (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, Master's thesis, Instituto de Estadística Aplicada y Computación, ULA, Mérida. (Disponible en <http://tesis.saber.ula.ve/theses/available/etd-06202003-171618/>).
- Burton, P. R. & Walls, J. (1987), 'Selection-adjusted comparison of life-expectancy of patients on continuous ambulatory peritoneal dialysis, haemodialysis and renal transplantation', *The Lancet* **i**, 1115–1119.
- Cox, D. (1972), 'Regression models and life tables (with discussion).', *Journal of the Royal Statistical Society: Series B* (34), 187–220.

- Fleming, T. R. & Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, John Wiley & Sons, Inc., N.Y.
- Greenwood, M. (1926), 'The natural duration of cancer', *Reports on Public Health and Medical Subjects* (33), 1–26.
- Held, P. J., Port, F. K., Turrene, M.Ñ., Gaylin, D. S., Hamburger, R. J. & Wolfe, R. A. (1994), 'Continuous ambulatory peritoneal dialysis and hemodialysis: Comparison of patients with adjustment for comorbid conditions', *Kidney International* (45), 1163–1169.
- Hosmer, D. W. & Lemeshow, S. (1999), *Applied Survival Analysis: Regression Modeling of Time to Event Data*, John Wiley & Sons, Inc., N.Y.
- Hougaard, P. (1995), 'Frailty models for survival data', *Lifetime Data Analysis* (1), 255–273.
- Hutchinson, T. A., Thomas, D. C., Lemieux, J. C. & Harvey, C. E. (1984), 'Prognostically controlled comparison of dialysis and renal transplantation', *Kidney International* (26), 44–51.
- Insightful Corporation, I. (2001), *S-PLUS 6 for Windows Guide to Statistics, Volume 2*, WA: Insightful Corporation, Inc., Seattle.
- Iseki, K., Miyasato, F., Tokuyama, K., Nishime, K., Ueshara, H., Shiohira, Y., Sunagawa, H., Yoshihara, K., Yoshi, S., Toma, S., Kowatari, T., Wake, T., Oura, T. & Fukiyama, K. (1997), 'Low diastolic blood pressure, hypoalbuminemia, and risk of death in a cohort of hemodialysis patients', *Kidney International* (51), 1212–1217.
- Kaplan, E. & Meier, P. (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* (53), 457–481.
- Liang, K., Self, S., Bandeen-Roche, K. & Zeger, Z. (1995), 'Some recent developments for regression analysis of multivariate failure time data', *Lifetime Data Analysis* (1), 403–415.
- Maggiore, Q., Nigrelli, S., Cicarelli, C., Grimaldi, C., Rossi, G. & Michelassi, C. (1996), 'Nutritional and prognostic correlates of bioimpedance indexes in hemodialysis patients', *Kidney International* (50), 2103–2108.
- Maiorca, R., Vonesh, E., Cancarini, G., Cantaluppi, A., Manili, L., Brunori, G., Camerini, C., Sèller, P. & Strada, A. (1988), 'A six-years comparison of patients and technique survivals in capd and hd', *Kidney International* (34), 518–524.
- Maiorca, R., Vonesh, E., Cavalli, P., De Vecchi, A., Giangrande, A., La Greca, G., Scarpione, L., Bragantini, L., Cancarini, G., Cantaluppi, A., Castelnovo, C., A., C., Poisetti, P. & Viglino, G. (1991), 'A multicenter selection-adjusted comparison of patients and technique survival on capd and hemodialysis', *Peritoneal Dialysis International* (11), 118–127.



- R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, Vienna, Austria.
- Therneau, T. & Grambsch, P. (2000), *Modeling Survival Data: Extending the Cox Model*, Springer-Verlag, N.Y.
- Therneau, T., Grambsch, P. & Fleming, T. (1990), 'Martingale-based residuals for survival models', *Biometrika* (77), 147–160.
- Vaupel, J., Manton, K. & Stallard, E. (1979), 'The impact of heterogeneity in individual frailty on the dynamics of mortality', *Demography* (16), 439–454.
- Woods, J., Port, F., Orzoul, S., Buoncristiani, U., Y. E., Wolfe, R. & Held, P. (1998), 'Clinical and biochemical correlates of starting daily hemodialysis', *Kidney International* (55), 2467–2476.