

---

## Comparación de métodos de estimación en regresión de Cox

### Comparison of Methods of Estimation in Regression of Cox

Javier Ramírez<sup>a</sup>  
jaramirez@correo.unicordoba.edu.co

Ever Regino<sup>b</sup>  
enrique-20-10@hotmail.com

Stalyn Guerrero<sup>c</sup>  
syguerrerog@unal.edu.co

---

#### Resumen

En este trabajo se comparan mediante simulación los métodos de estimación Breslow, Efron y Exacto en Regresión de Cox, para encontrar la estimación de los parámetros del modelo, obteniendo intervalos de confianza mediante remuestreo *Bootstrap*, *Jackknife* y tradicionales asintóticos. Se generan muestras de tiempos utilizando la transformación inversa, para modelos de regresión Exponencial y Weibull. Se ilustran los resultados de las amplitudes de los intervalos de confianza tomando como referencia la estimación de la regresión paramétrica. Mostrando la eficiencia de dichos intervalos.

**Palabras clave:** Regresión de Cox, *Bootstrap*, *Jackknife*.

#### Abstract

In this work are compared using simulation methods for estimating Breslow, Efron and exact in Cox regression, to find the estimate of the model parameters, obtaining confidence intervals by resampling the *Bootstrap*, *Jackknife* and traditional asymptotic. Are generated samples of times using the inverse of the transformation, for models of exponential regression and Weibull. It illustrates the results of the amplitudes of the confidence intervals taking as a reference the regression estimate parametric. Showing the efficiency of these intervals.

**Keywords:** Regression of Cox, *Bootstrap*, *Jackknife*.

---

<sup>a</sup>Profesor Asociado. Departamento de Matemáticas y Estadística. Universidad de Córdoba, Colombia

<sup>b</sup>Estadístico (c). Departamento de Matemáticas y Estadística. Facultad de Ciencias Básicas. Universidad de Córdoba. Montería. Colombia

<sup>c</sup>M.Sc (c) Estadística. Departamento de Estadística. Facultad Ciencias. Universidad Nacional de Colombia. Bogotá D.C., Colombia

## 1. Introducción

En análisis de Confiabilidad, tratado en la Industria o Supervivencia en las Ciencias Médicas, modelar el tiempo de ocurrencia de un evento de interés involucra en muchos casos efectos de covariables, donde la regresión paramétrica de análisis de supervivencia se utiliza cuando los tiempos de vida se ajustan a modelos probabilísticos conocidos. Utilizando la función `survreg` de R; sin embargo, en la actualidad diversas áreas no hacen de manera intensiva las pruebas de bondad de ajuste como también los tiempos no se ajustan de manera significativa a los modelos paramétricos, donde la regresión de Cox resulta una estrategia muy usada (Kalbfleisch & Prentice 1980), teniendo cuidado que en el caso univariado sin covariables la estimación no paramétrica tienen muy buenos resultados en algunas condiciones (Ramírez & Martínez 2015). En la sección 2 se presenta una descripción del modelo de regresión de Cox y los elementos que lo componen, luego en la sección 3 se muestran los diferentes métodos de estimación clásicos para encontrar la estimación de los parámetros del modelo.

Finalmente en la sección 4 se realiza un estudio vía simulación, utilizando los métodos de Breslow, Efron y Exacto, planteando escenarios de simulación, con tamaños de muestra de  $n = (25, 50, 100)$  a partir de una simulación de MonteCarlo, generando los tiempos de falla, mediante el método de la transformación inversa para regresión de Cox (Bender et al. 2003), en la literatura no se encuentran trabajos donde se comparen mediante simulación estos métodos simultáneamente; sin embargo, algunos autores utilizan las estrategias de intervalos de confianza asintóticos, *Bootstrap* y *Jackknife* en la comparación de métodos de estimación, (Ramírez et al. 2015), presentando buenos resultados en tamaños de muestra pequeños con remuestreo *Jackknife*. Adicionalmente, en la sección 5 se presentan las conclusiones.

## 2. Modelo de regresión de Cox

Una de las técnicas estadísticas más utilizadas para evaluar la relación entre un conjunto de variables explicativas y el tiempo de supervivencia es el modelo de regresión de riesgos proporcionales conocido también como modelo de *Cox* (Cox 1972).

$$h(t; \mathbb{X}) = h_0(t)e^{\beta\mathbb{X}} = h_0(t)e^{(\beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p)} \quad (1)$$

Este modelo está compuesto por el producto de dos términos. El primero depende exclusivamente del tiempo, mientras que el segundo depende solamente de las variables explicativas  $\mathbb{X}$ .

- $h(t; \mathbb{X})$  : Tasa de riesgo de un sujeto, con valores  $\mathbb{X} = (x_1, x_2, \dots, x_p)$  en las variables explicativas en el instante  $t$ . Es la variable respuesta que se

modeliza. Representa el riesgo de fallecer en el instante  $t$  de los sujetos que tienen un determinado patrón de valores  $x$  en las variables explicativas.

- $e^{\beta' \mathbb{X}}$  : función exponencial, cuyo exponente es la combinación lineal, sin término constante, de las  $p$  variables explicativas  $x_i$ .
- $h_0(t)$  : función de riesgo de referencia (“*baseline*” o “*underlying hazard function*”), que sólo depende del tiempo, llamada así por que representa las tasas instantánea de riesgo de un sujeto hipotético con valor 0 en todas las variables predictivas (ya que el término exponencial  $e^0 = 1$ ).

El modelo de *Cox*, producto de estas dos funciones, se caracteriza por no especificar la forma de  $h_0(t)$ . Se denomina un modelo semiparamétrico porque se estiman los  $p$  parámetros  $\beta_i$  mientras que el valor de la función  $h_0(t)$  se obtiene a través de los datos.

Observe su similitud con el modelo de regresión logística expresado en forma de razón de chance de sujetos con “respuesta” presente ( $Y = 1$ ).

$$\frac{\pi_x}{1 - \pi_x} = e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} = e^{\beta_0} e^{(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \quad (2)$$

Puede comprobar que el término constante  $e^{\beta_0}$  de esta razón de chance se sustituye en el modelo de *Cox* por la función de riesgo de referencia  $h_0(t)$  que no depende de las variables explicativas  $\mathbb{X}$ . Es fácil expresar el modelo de *Cox* en función de la tasa de riesgo acumulado  $H(t; x)$  o de la probabilidad acumulada de supervivencia  $S(t; x)$ , a través de la estrecha relación existente entre las distribuciones.

- $h(t; x)$  : Tasa instantánea de riesgo. Representa el riesgo que un sujeto (con valores  $\mathbb{X}$  en las variables explicativas) tiene de fallecer en el instante  $t$ .
- $H(t; x)$  : Tasa acumulada de riesgo. Representa el riesgo que un sujeto (con valores  $\mathbb{X}$ ) tiene de fallecer hasta el tiempo  $t$

$$H(t; x) = H(t) e^{\beta' \mathbb{X}} \quad (3)$$

- $S(t; x)$ : Función de supervivencia. Representa la probabilidad de supervivencia en el tiempo  $t$  de los sujetos con un determinado patrón de valores  $\mathbb{X}$  en las variables explicativas.

se calcula a través de su relación con la función de riesgo acumulado

$$S(t; x) = e^{-H(t; x)} \quad (4)$$

Tomando logaritmo natural, la tasa de riesgo acumulado viene dada por:

$$H(t; x) = -\ln S(t; x) \quad (5)$$

Se tiene que las funciones se establecen una en función de las demás (Kalbfleisch & Prentice 1980)

En análisis de sobrevivencia ocurre a menudo el fenómeno de la censura, esto es, la no determinación del tiempo exacto del evento de interés, suministrando los datos información parcial de los tiempos. Existen diferentes tipos de censura, utilizando en este estudio censura a derecha. Por lo general ocurre cuando al finalizar el experimento no ocurre el evento de interés.

### 3. Métodos de estimación en regresión de Cox

Sea  $t_1 < t_2 < \dots < t_d$  denota los tiempos distintos de ocurrencia del evento. Sea  $d_i$  el número de fallas en  $t_i$  y  $D_i$  el conjunto de etiqueta de los individuos que presentan el evento en el tiempo  $t_i$ . Sea  $S_i$  la suma de los vectores  $X_k$  sobre todos los individuos que presentan el evento en  $t_i$ . Esto es  $S_i = \sum_{k \in D_i} X_k$ .

Sea  $R_i$  el conjunto de todos los individuos en riesgo justo antes de  $t_i$ , (Meeker & Escobar. 1998). La verosimilitud básica es:

$$L(\beta) = \frac{\exp(x_i' \beta)}{\sum_{l \in R_i} \exp(x_l' \beta)}$$

cuando los datos contienen tiempos observados empatados la verosimilitud parcial puede llevar un tiempo de computación considerable. Por esta razón, cuando se tienen datos con empates, se utilizan aproximaciones para la función de verosimilitud parcial.

Hay varias propuestas para para construir la verosimilitud parcial cuando hay empates entre los tiempos de los eventos, propuestos por Breslow (1974), Efron (1982) y Cox (1972).

#### 3.1. Método de Breslow

Supongamos  $k$  tiempos de falla para  $n$  individuos no censurados y  $n - k$  censurados a derecha. Sean  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  tiempos ordenados distintos con sus correspondientes covariables  $x_1, x_2, \dots, x_k$ . Una aproximación de la verosimilitud parcial bastante utilizada fue sugerida por Breslow (1974). Esta aproximación considera que los  $d_i$  eventos al tiempo  $t_i$  son distintos y ocurren secuencialmente. La aproximación viene dada por la siguiente fórmula

$$L(\beta) = \prod_{i=1}^d \frac{\exp(\beta^t S_i)}{\left[ \sum_{k \in R_i} \exp(\beta^t X_k) \right]^{d_i}} \quad (6)$$

y el logaritmo de la verosimilitud

$$\log(L_I(\beta)) = \sum_{i=1}^d \left[ \beta^t S_i - d_i \log \left( \sum_{k \in R_i} \exp(\beta^t X_k) \right) \right] \quad (7)$$

entonces para estimar los parámetros del modelo tenemos

$$\frac{\partial \log(L_I(\beta))}{\partial \beta} = \sum_{i=1}^d \left[ S_i - \frac{d_i \sum_{k \in R_i} X_k \exp(\beta^t X_k)}{\sum_{k \in R_i} \exp(\beta^t X_k)} \right] = 0 \quad (8)$$

Obteniéndose estos resultados a través del método numérico de Newton-Raphson. Suponiendo  $\beta$  parámetro de interés con  $\dim(\beta) = p$ , entonces los pasos son:

1.  $k = 1$
2. Se encuentra  $\beta^{(k)}$
3. Resolvemos  $I(\beta^{(k)})(\beta^{(k+1)} - \beta^{(k)}) = U(\beta^{(k)})$  para  $\beta^{(k+1)}$
4. Incrementamos  $k$  por 1.
5. Volver al paso 3 y repetir hasta la convergencia.

donde  $U(\beta) = \left( \frac{\partial I(\beta)}{\partial \beta_1} \right)'$  y  $I(\beta) = - \left( \frac{\partial^2 I(\beta)}{\partial \beta_r \partial \beta_s} \right)$

### 3.2. Método de Efron

Una aproximación alternativa, sugerida por Efron (1977), la cual se considera más exacta que la de Breslow, es

$$L_{II}(\beta) = \prod_{i=1}^d \frac{\exp(\beta^t X_k)}{\prod_{j=1}^{d_i} \left[ \sum_{k \in R_i} \exp(\beta^t X_k) - \frac{j-1}{d_j} \sum_{k \in D_i} \exp(\beta^t X_k) \right]} \quad (9)$$

y el logaritmo de la verosimilitud

$$\log(L_{II}(\beta)) = \sum_{i=1}^d \left[ \beta^t X_k - \sum_{j=1}^{d_i} \log \left( \sum_{k \in R_i} \exp(\beta^t X_k) - \frac{j-1}{d_j} \sum_{k \in D_i} \exp(\beta^t X_k) \right) \right] \quad (10)$$

entonces para estimar los parámetros del modelo tenemos

$$\frac{\partial(\log(L_{II}(\beta)))}{\partial \beta} = \sum_{i=1}^d \left[ X_k - \sum_{j=1}^{d_i} \frac{\sum_{k \in R_i} X_k \exp(\beta^t X_k) - \frac{j-1}{d_j} \sum_{k \in D_i} X_k \exp(\beta^t X_k)}{\sum_{k \in R_i} \exp(\beta^t X_k) - \frac{j-1}{d_j} \sum_{k \in D_i} \exp(\beta^t X_k)} \right] = 0 \quad (11)$$

Los cuales se hace vía métodos numéricos. De forma análoga como en el método anterior.

### 3.3. Método de la verosimilitud parcial exacta

La verosimilitud se construye asumiendo un modelo logístico para la tasa de riesgo (Kalbfleisch & Prentice 1980), esto es, si tenemos  $h(t|x)$  denota la probabilidad condicional de fallo en el intervalo  $(t, t+1)$  dada la sobrevivencia al inicio del intervalo y si asumimos un modelo de tasa de riesgo

$$\frac{h(t|x)}{1-h(t|x)} = \frac{h_0(t)}{1-h_0(t)} \exp(\beta^t X) \quad (12)$$

Para construir la verosimilitud, sea  $Q_i$  que denota el conjunto de todos los subgrupos de individuos  $d_i$  que podrían ser seleccionados frente al riesgo establecido  $R_i$ . Cada elemento de  $Q_i$  es una  $d_i$ -tupla que podría ser uno de los fallos  $d_i$  al tiempo  $t_i$ . Sea  $q = (q_1, \dots, q_{d_i})$  uno de estos elementos de  $Q_i$  y se define  $S_q^* = \sum_{k=1}^{d_i} X_{qk}$  entonces la función de verosimilitud está dada por

$$L_{III}(\beta) = \prod_{i=1}^d \frac{\exp(\beta^t X_k S_i)}{\sum_{q \in Q_i} \exp(\beta^t S_q^*)} \quad (13)$$

$$\log(L_{III}(\beta)) = \sum_{i=1}^d \left[ \beta^t X_k S_i - \log \left( \sum_{q \in Q_i} \exp(\beta^t S_q^*) \right) \right] \quad (14)$$

entonces para estimar los parámetros del modelo tenemos

$$\frac{\partial \log(L_{III}(\beta))}{\partial \beta} = \sum_{i=1}^d \left[ X_k S_i - \frac{\sum_{q \in Q_i} S_q^* \exp(\beta^t S_q^*)}{\sum_{q \in Q_i} \exp(\beta^t S_q^*)} \right] = 0 \quad (15)$$

Los cuales se hacen por métodos numéricos descritos luego de la ecuación (11).

### 3.4. Estimación mediante métodos de remuestreo

Se trata de una técnica de simulación basada en la información de interés. Su mayor ventaja es que reduce la necesidad de asumir determinado modelo probabilístico para las observaciones. En general, requiere un número elevado de cálculos.

#### 3.4.1. Método de remuestreo *Bootstrap*

Suponga una muestra  $X = (x_1, x_2, \dots, x_n)$  en donde  $x_i$  se extrae de una distribución empírica  $\tilde{F}$  (o de una población). Muestras de tamaño  $n$  son extraídas de  $x$  con remplazo. Hay  $n^n$  posibles muestras, llamadas Muestras *Bootstrap*. La estimación *Bootstrap* del error estándar es la desviación estándar de las repliaciones *Bootstrap*:

$$\widehat{SE}_{boot} = \left\{ \sum_{b=1}^B \frac{[\beta(x^{*b}) - \beta(\cdot)]^2}{(B-1)} \right\}^{\frac{1}{2}} \quad (16)$$

donde

$$\beta(\cdot) = \frac{1}{B} \sum_{b=1}^B \beta(x^{*b}) \quad (17)$$

corresponde al promedio de las remuestras bootstrap.

#### 3.4.2. Método de remuestreo *Jackknife*

Una de las primeras técnicas para obtener los estimadores estadísticos fiables es la técnica *Jackknife*. Se requiere menos potencia de cálculo que las técnicas más

recientes. Supongamos que tenemos una muestra  $X = (x_1, x_2, \dots, x_n)$  y un estimador  $\hat{\theta} = \hat{\beta}_{(i)}$ . La técnica *Jackknife* se centra en las muestras que dejan de lado a una observación a la vez:

$$X_i = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (18)$$

Para  $i = 1, 2, \dots, n$  estas son llamadas muestras *Jackknife*. La  $i$ -ésima muestra *Jackknife* consiste en el conjunto de datos eliminando la  $i$ -ésima observación. Sea  $\hat{\theta}_i = \hat{\beta}_{(i)}$  la  $i$ -ésima replicación *Jackknife* de  $\hat{\theta}$

La estimación del error estándar *Jackknife* se define por:

$$\widehat{SE}_{Jack} = \left[ \frac{n-1}{n} \sum_{i=1}^n \left( \hat{\beta}_i - \hat{\beta}_{(\cdot)} \right)^2 \right]^{\frac{1}{2}} \quad (19)$$

Donde

$$\hat{\beta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_{(i)} \quad (20)$$

corresponde al promedio de los pseudo valores *Jackknife*.

## 4. Intervalos de confianza utilizados

Para comparar la estimación de los parámetros del modelo de regresión de Cox se utilizan intervalos de confianza asintóticos de Anderson, *Bootstrap* y *Jackknife*:

### 4.1. Intervalos de confianza de Anderson

Sean  $\beta_1 > \beta_2 > \dots > \beta_p$  los parámetros del modelo de regresión de Cox. Sean  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  los estimadores respectivamente. Un intervalo de confianza de  $100(1 - \alpha)\%$  para los parámetros del modelo de regresión de Cox  $\beta_i$  son:

$$\left[ \frac{\hat{\beta}_i}{1 + \sqrt{\frac{2}{n}} Z_{\alpha/2}}, \frac{\hat{\beta}_i}{1 - \sqrt{\frac{2}{n}} Z_{\alpha/2}} \right] \quad (21)$$

con  $i = 1, 2, \dots, p$  y  $Z_{\alpha/2}$  cuantil de la distribución normal.

### 4.2. Intervalos de confianza *Bootstrap*

Un intervalo de confianza para los parámetros de modelo de regresión de Cox está dado por:

$$\left[ \hat{\beta}^* - Z_{\alpha/2} \sigma_{\beta}^*, \hat{\beta}^* + Z_{\alpha/2} \sigma_{\beta}^* \right] \quad (22)$$

donde  $\sigma_{\beta}^*$  corresponde al error estándar *Bootstrap* dado en la ecuación (16) y  $\beta^*$  en teniendo en cuenta el método de estimación Breslow, Efron y Exacto.



### 4.3. Intervalo de confianza *Jackknife*

Un intervalo de confianza para los parámetros del modelo de regresión de Cox mediante la técnica *Jackknife* está dado por:

$$\left[ \widehat{\beta}_{n-1,l} - Z_{\alpha/2} \sqrt{v_{jk}}, \widehat{\beta}_{n-1,l} + Z_{\alpha/2} \sqrt{v_{jk}} \right] \quad (23)$$

donde el error estándar *Jackknife* esta dado en (19) y  $\widehat{\beta}$  en (20).

## 5. Escenarios de simulación

Se generan 2000 muestras artificiales mediante el método de la transformación inversa de la función de distribución acumulada con el software **R** (R Development Core Team 2008), tomando tres parámetros en los modelos generados con dos covariables. Una  $N(0, 1)$  y otra producto de  $N(0, 1)$  por  $Beta(\alpha = 2, 4, 6)$ , mediante tiempos generados *Exponencial* ( $\lambda = 5$ ) y *Weibull* ( $\alpha = 5, \lambda = 2$ ), muy utilizados en sobrevivencia (Ramírez et al. 2016), con porcentajes de censura (0 %, 15 %, 30 %) y tamaños de muestra ( $n = 25, 50, 100$ ), comparando los métodos de estimación Breslow, Efron y Exacta. La función de sobrevivencia del modelo de hazard proporcional de Cox es dado por:

$$S(t|x) = \exp[-H_0(t) \exp(\beta'x)] \quad (24)$$

donde

$$H_0(t) = \int_0^t h_0(u) du \quad (25)$$

Corresponde a la función baseline, entonces la función de distribución acumulada del modelo de Cox es:

$$F(t|x) = 1 - \exp[-H_0(t) \exp(\beta'x)] \quad (26)$$

Para una variable aleatoria  $Y$  con distribución  $F$ , entonces  $U = F(Y)$  sigue una distribución uniforme, es decir  $U[0, 1]$ , en efecto  $(1 - U) \sim Uni[0, 1]$ , entonces:

$$U = \exp[-H_0(t) \exp(\beta'x)] \quad (27)$$

Por lo tanto

$$T = H_0(t)^{-1}[-\log(U) \exp(\beta'x)] \quad (28)$$

Bender et al. (2003) proponen para el caso exponencial con parámetro de escala  $\lambda > 0$ , tiene como expresión para los tiempos de falla dados por:

$$T = -\frac{\log U}{\lambda \exp(\beta'x)} \quad (29)$$

Para el caso Weibull con parámetro de escala  $\lambda$  y de forma  $\nu$  se obtienen los tiempos de falla generando

$$T = - \left[ \frac{\log U}{\lambda \exp(\beta^{\nu} x)} \right]^{1/\nu} \quad (30)$$

Luego los resultados se presentan a continuación.

### 5.1. Resultados Simulación

Utilizando 2000 muestras simuladas y 500 remuestras *Bootstrap*, son calculados los intervalos de confianza asintóticos, *Bootstrap* y *Jackknife*, graficando las amplitudes promedio de los intervalos para cada método de estimación (Breslow, Efron y Exacta) y parámetros del modelo de Regresión de Cox ( $\beta_1, \beta_2, \beta_3$ ), con los diferentes tamaños de muestra ( $n = 25, 50, 100$ ) y porcentajes de censura (0%, 15%, 30%).

Utilizando las siguientes siglas;  $BX_i = \hat{\beta}_i$ , con  $i = 1, 2, 3$  correspondiente a la estimación mediante el método Breslow para cada uno de los tres parámetros del modelo.  $EFX_i = \hat{\beta}_i$ , con  $i = 1, 2, 3$  representa la estimación mediante el método Efron para cada uno de los parámetros.  $EX_i = \hat{\beta}_i$  corresponde a la estimación utilizando el método Exacto. Finalmente  $PX_i = \hat{\beta}_i$  es la estimación paramétrica de los tres parámetros del modelo. Es de resaltar que AN representa las amplitudes de los intervalos asintóticos, AB de los intervalos con remuestreo *Bootstrap* y AJ las amplitudes promedio de los intervalos *Jackknife*.

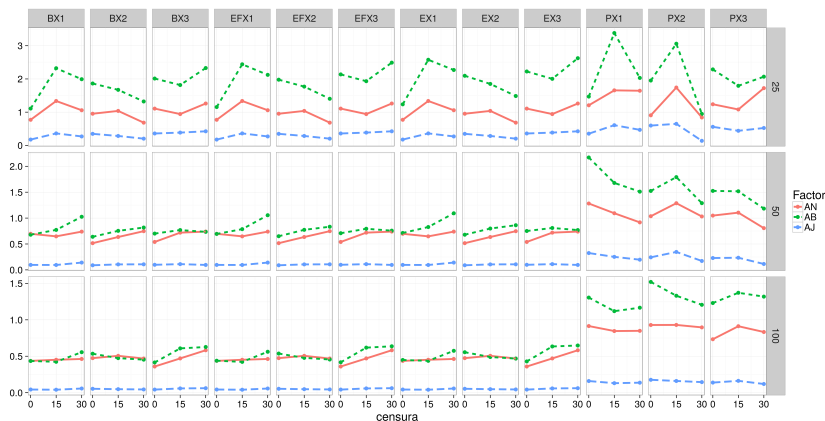


Figura 1: Amplitudes de los intervalos de confianza (caso Exponencial). Fuente: elaboración propia.

Notando que los intervalos de confianza mediante el remuestreo *Jackknife* presentan mejores resultados; es decir, menores amplitudes promedio, teniendo en

cuenta el parámetro estimado del estudio, tamaño de muestra y porcentaje de censura. Seguido de los intervalos asintóticos y finalmente bootstrap, resaltando que los dos últimos tienen resultados similares cuando se aumenta el tamaño de muestra. Por otra parte las amplitudes de los intervalos de confianza para los parámetros del modelo de regresión de Cox, con los métodos de estimación Breslow, Efron y Exacta resultan ser muy similares, presentando pequeñas diferencias cuando se aumenta el porcentaje de censura a 30 %, resaltando que el intervalo de confianza paramétrico posee mayor amplitud en los intervalos (ver figura 1)

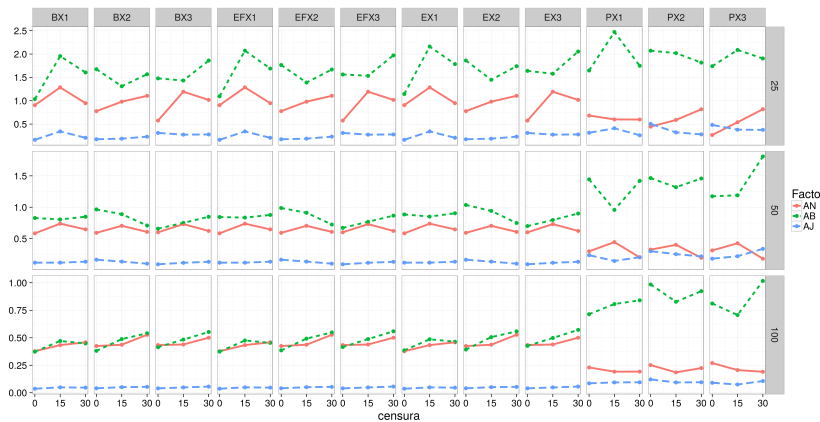


Figura 2: Amplitudes de los intervalos de confianza (caso Weibull). Fuente: elaboración propia.

De igual forma se manifiestan los comportamientos bajo la distribución Weibull en los tiempos de falla en la regresión de Cox con tres parámetros, resaltando que en muestras pequeñas el remuestreo *Jackknife* presenta mejores amplitudes de los intervalos de confianza con nivel de cobertura real de simulación similar al nominal del 95 %. Es de resaltar que los intervalos en general presentaron mayores coberturas reales de simulación frente a las nominales del 95 %, tomando como referencias a las estimaciones paramétricas de *Survreg* de R, y centrando en este trabajo solo en las amplitudes de los mismos, reflejado por intervalos más amplios. Se opta por mostrar solo las amplitudes. (Ver figura 2)

## 6. Conclusiones y recomendaciones

El hecho de que la muestra *Jackknife* delete I, elimine consecutivamente un elemento tiende a parecerse a la muestra original y cuya variabilidad de los pseudovalores *Jackknife* posean una varianza menor, ya que en el caso del remuestreo *Bootstrap* tomar muestras con reemplazamiento de la muestra original hace que se presenten replicaciones de tiempos ya sean de falla o censurados en las muestras afectando la estimación y corrigiéndose a muestras muy grandes.

Para establecer diferencias mayores entre los métodos de estimación en regresión de Cox, se deben tener situaciones más extremas de altos porcentajes de censura y bajos tamaños de muestra. Ya que con los escenarios utilizados no se presentan diferencias entre las estimaciones.

**Recibido: 1 de junio de 2016**

**Aceptado: 10 de noviembre de 2016**

## Referencias

- Bender, R., Augustin, T. & Blettner, M. (2003), 'Generating survival time to simulate cox proportional hazards models.'
- Breslow, N. (1974), 'Covariance Analysis of Survival Data under the Proportional Hazards Model.', *International Statistical Review* **43**, 43–54.
- Cox, D. R. (1972), 'Regression Models and Life Tables (with Discussion)', *Journal of the Royal Statistical Society B* **34**, 187–220.
- Efron, B. (1977), 'The Efficiency of Coxs Likelihood Function for Censored Data.', *Journal of the American Statistical Association* **72**, 557–565.
- Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics.
- Kalbfleisch, J. & Prentice, R. (1980), *The statistical analysis of failure time data*, New York, Wiley.
- Meeker, W. & Escobar., L. (1998), *Statistical Models and Methods for Reliability Data*, New York, John Wiley and Sons.
- R Development Core Team (2008), *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>
- Ramírez, J., I. Osuna, J. R. & Guerrero, S. (2016), 'Remuestreo Bootstrap y Jackknife en confiabilidad: Caso Exponencial y Weibull', *Revista Facultad de Ingeniería* **25**(41), 55–62.
- Ramírez, J. & Martínez, G. (2015), 'Comparación de Estimadores no paramétricos frente a los paramétricos para la función de Confiabilidad', *Revista Ingeniería Investigación y Desarrollo* **15**(2), 15–24.
- Ramírez, J., Martínez, G. & Guerrero, S. (2015), 'Intervalos de confianza para valores propios en el análisis de correspondencias a partir de una muestra probabilística', *Revista Comunicaciones en Estadística* **8**(2), 225–243.