

**Comparación de Intervalos de Confianza para la
Función de Supervivencia, en un tiempo de interés,
con Censura a Derecha**

Javier Ramírez Montoya

Director
Eva Cristina Manotas, Ms.C

Trabajo Presentado Como Requisito Para Optar al Título de Magíster En Estadística

Escuela de Estadística
Universidad Nacional de Colombia
Sede Medellín

Medellín, Antioquia
Junio 2010

Dedicatoria

A Dios, a mis padres por su comprensión.

Agradecimientos

El autor del presente trabajo de grado agradece de manera muy especial a todas aquellas personas que contribuyeron en la elaboración y corrección del mismo. En particular agradece a la Escuela de Estadística, a la profesora Eva Cristina M. por sus valiosas sugerencias y a los jurados.

Resumen

En este trabajo se comparan métodos para encontrar intervalos de confianza para la función de supervivencia, como los métodos de remuestreo *Bootstrap* aplicado a los estimadores de Kaplan-Meier y Nelson-Aalen. También mediante las transformaciones \log , $\log(-\log)$ y *Arco seno* que pueden resultar en muchos casos más efectivos. Además se determina la eficiencia que presentan los intervalos de confianza no paramétricos frente a los paramétricos.

Palabras Clave: Función de supervivencia, Intervalos de confianza, Censura a derecha, *Bootstrap*, Estimador Kaplan-Meier, Estimador Nelson-Aalen, Transformaciones \log , $\log(-\log)$ y *Arcsen*.

Abstract

In this work, its compared simulation alternative methods to find confidence intervals for the survival function such as the resampling methods *Bootstrap*, aplicated to the Kaplan-Meier and Nelson-Aalen estimators. Also through \log , $\log(-\log)$ and *Arcsen* transformations that can result more efectives in many cases. It also determines the efficiency presented by the nonparametric confidence intervals compared with parametric.

Key words: Survival function, Confidence intervals, Right censoring, Bootstrap, Kaplan-Meier estimator, Nelson-Aalen estimator, \log , $\log(-\log)$ and *Arcsen* Transformations.

Tabla de Contenido

1. MARCO TEÓRICO	1
1.1. Funciones de Básicas	4
1.1.1. Censura	6
1.1.2. Tipos de Censura	6
1.2. Intervalos de Confianza paramétricos	8
1.2.1. Caso Exponencial:	8
1.2.2. Caso Weibull:	9
1.3. Estimadores no paramétricos	10
1.3.1. Estimador de Kaplan Meier Para la Función De Supervivencia .	10
1.3.2. Estimador de Nelson-Aalen Para la Función De Supervivencia .	13
1.4. Intervalos de confianza no paramétricos	14
1.4.1. Intervalo de Confianza mediante el estimador de Kaplan Meier .	14
1.4.2. Intervalo de Confianza mediante el estimador de Nelson-Aalen .	15
1.4.3. Intervalos de Confianza Mediante Transformaciones	15
1.4.4. El Método de Remuestreo Bootstrap Aplicado a datos de Super- vivencia	16
1.4.5. Criterio de Comparación de los Intervalos de Confianza	18
2. ESTUDIO DE SIMULACIÓN	19

2.1. Escenarios de Simulación	19
2.2. Resultados	20
2.2.1. Resultados para $n = 25$	20
2.2.2. Resultados para $n = 50$	22
2.2.3. Resultados para $n = 75$	24
2.2.4. Resultados para $n = 100$	26
3. CONCLUSIONES	31
4. REFERENCIAS	32
5. ANEXOS	35
5.1. Graficos	35
5.2. Documentación Funciones y Códigos en R	46
5.2.1. Código	48

Índice de Gráficos

5.1. Tasas de Error de I.C para $S(t)$ con $n = 25$	36
5.2. Tasas de Error de I.C para $S(t)$ con $n = 25$	37
5.3. Tasas de Error de I.C para $S(t)$ con $n = 25$ y $n = 50$	38
5.4. Tasas de Error de I.C para $S(t)$ con $n = 50$	39
5.5. Tasas de Error de I.C para $S(t)$ con $n = 50$	40
5.6. Tasas de Error de I.C para $S(t)$ con $n = 75$	41
5.7. Tasas de Error de I.C para $S(t)$ con $n = 75$	42
5.8. Tasas de Error de I.C para $S(t)$ con $n = 75$ y $n = 100$	43
5.9. Tasas de Error de I.C para $S(t)$ con $n = 100$	44
5.10. Tasas de Error de I.C para $S(t)$ con $n = 100$	45

Índice de Tablas

2.1. TE y LPI de NN 95 % con 0 % de censura para $n = 25$	20
2.2. TE y LPI de NN 95 % con 15 % de censura para $n = 25$	20
2.3. TE y LPI de NN 95 % con 25 % de censura para $n = 25$	21
2.4. TE y LPI de NN 95 % con 35 % de censura para $n = 25$	21
2.5. TE y LPI de NN 95 % con 45 % de censura para $n = 25$	22
2.6. TE y LPI de NN 95 % con 0 % de censura para $n = 50$	22
2.7. TE y LPI de NN 95 % con 15 % de censura para $n = 50$	22
2.8. TE y LPI de NN 95 % con 25 % de censura para $n = 50$	23
2.9. TE y LPI de NN 95 % con 35 % de censura para $n = 50$	23
2.10. TE y LPI de NN 95 % con 45 % de censura para $n = 50$	24
2.11. TE y LPI de NN 95 % con 0 % de censura para $n = 75$	24
2.12. TE y LPI de NN 95 % con 15 % de censura para $n = 75$	24
2.13. TE y LPI de NN 95 % con 25 % de censura para $n = 75$	25
2.14. TE y LPI de NN 95 % con 35 % de censura para $n = 75$	25
2.15. TE y LPI de NN 95 % con 45 % de censura para $n = 75$	25
2.16. TE y LPI de NN 95 % con 0 % de censura para $n = 100$	26
2.17. TE y LPI de NN 95 % con 15 % de censura para $n = 100$	26
2.18. TE y LPI de NN 95 % con 25 % de censura para $n = 100$	27

2.19. TE y LPI de NN 95 % con 35 % de censura para $n = 100$	27
2.20. TE y LPI de NN 95 % con 45 % de censura para $n = 100$	27
2.21. Índice de IC para $\hat{S}_{KM}(t)$ con 15 % de censura	29
2.22. Índice de IC para $\hat{S}_{NA}(t)$ con 15 % de censura	30

INTRODUCCIÓN

En la actualidad existen diferentes métodos para encontrar intervalos de confianza para la función de supervivencia en un tiempo de interés, con censura a derecha, tales como los métodos tradicionales utilizando los estimadores de Kaplan-Meier y Nelson-Aalen, también mediante remuestreo *Bootstrap* aplicado a estos estimadores y a través de las transformaciones \log , $\log(-\log)$ y *Arco seno*, que pueden resultar en muchos casos más efectivos. Este trabajo proporciona criterios para establecer que intervalos de confianza no paramétricos utilizar para estimar la función de supervivencia en un tiempo de interés, con censura a derecha y determinar la eficiencia de los estimadores no paramétricos frente a los paramétricos.

En el capítulo 1 se muestra inicialmente una revisión de los artículos relacionados con la comparación de intervalos de confianza para la función de supervivencia en tiempos de interés y luego los conceptos fundamentales en análisis de supervivencia para la construcción de intervalos de confianza utilizando los estimadores de Kaplan-Meier y Nelson-Aalen, mediante las diferentes transformaciones de los mismos, como son la transformación \log , $\log(-\log)$ y *Arco seno*. En el capítulo 2 se encuentra la descripción de los escenarios de simulación y del procedimiento implementado en R, así como sus resultados y conclusiones en el capítulo 3, con base en los criterios de comparación. En el capítulo 4 se presentan las referencias utilizadas en este trabajo.

Al final del documento, en los anexos, se encuentra la documentación completa de las funciones creadas en R para la comparación de los intervalos de confianza para la función de supervivencia en un tiempo de interés, con base en las distribuciones Weibull y Exponencial para los tiempos de falla/censura.

Capítulo 1

MARCO TEÓRICO

En el análisis de supervivencia por lo general se desea estimar la función de supervivencia, así, dicha estimación puede realizarse utilizando procedimientos paramétricos o no paramétricos, teniendo en cuenta el desconocimiento de la distribución de dicha variable, por lo que en el campo de la Medicina en muchos casos se utilizan estimadores no paramétricos, como son Kaplan-Meier y Nelson-Aalen cuando se presenta el fenómeno de censura a derecha, para luego estimar dicha función de supervivencia a través de un intervalo de confianza, donde se utilizan los intervalos de confianza tradicionales sin tener en cuenta que existen otros intervalos de confianza que pueden resultar en muchos casos más efectivos cuando se presenta el fenómeno de censura a derecha.

En éste trabajo se consultó el estado del arte asociado a los intervalos de confianza para la función de supervivencia, en tiempos de interés, cuando se presenta el fenómeno de censura a derecha en los datos y con el fin de hacer comparables los resultados de este trabajo con artículos realizados por diferentes autores, se encontraron artículos relacionados con tiempo de interés como el tiempo de supervivencia mediano, y algunos tiempos arbitrarios.

Jeng y Meeker (2000), comparan intervalos de confianza normales, transformados de la aproximación normal, razón de logverosimilitud, razón de logverosimilitud con corrección de Bartlett, Bootstrap-t paramétrico, transformación Bootstrap-t paramétrico, raíz de la razón logverosimilitud Bootstrap paramétrico, percentil Bootstrap paramétrico, Bootstrap paramétrico con sesgo corregido, usando gráficas, también muestra la distribución de $Z_{t_{0.1}}$, con 2000 simulaciones.

Barber y Jenninson (1999), introducen dos nuevas pruebas del valor de la función de la supervivencia, con 5000 simulaciones, calcularon mediante la fórmula de Greenwood, con $n = 100$ tiempos y distribuciones de censura exponenciales, también compara los intervalos de confianza mediante la transformación $\log(-\log)$ y aproximación Beta, Bootstrap y de razón de verosimilitud.

Borgan y Liestol (1990), comparan intervalos de confianza para la función de supervivencia tradicionales, transformados por $\log(-\log)$ y transformación de raíz de arcoseno, usando $n = 25, 50$ y 200 , para los tiempos de interés ($t_i = 0.2, 0.6$ y 1.0), con las tasas de errores usando distribución de tiempos/censura (Exponencial/Exponencial, Weibull/Exponencial), con 10000 réplicas.

Bie, Borgan y Liestol (1987), construyen intervalos de confianza para la función hazard acumulada, a través del estimador de Nelson Aalen, Kaplan Meier, transformación logarítmica y arcoseno, con tamaños de muestra $n = 25, 50$ y 200 , utilizando distribución de tiempos/censura (Exponencial/Exponencial, Exponencial/Uniforme y Weibull/Uniforme) en los tiempos de interés ($0.2, 0.6$ y 1.0), compararon los intervalos de confianza estándar, transformación logarítmica y arcoseno a través de la tasa de error con nivel de confianza del 95 %, también cambian los niveles de confianza de 90 %, 95 % y 99 % con 20000 simulaciones para el tiempo 0,4 y porcentajes de censura 0 %, 25 %, 50 % y 75 %.

Akritas (1986), Muestra mediante remuestreo bootstrap una comparación de bandas de confianza para la función de supervivencia con niveles $\alpha = 0.01, 0.05, 0.1$ y 0.2 , tamaños de muestra $n = 25$ y 50 y porcentajes de censura de 20 % y 40 % usando distribución de tiempos/censura exponencial/uniforme.

Jenninson y Turnbull (1985), comparan intervalos de confianza para el tiempo de supervivencia mediano, usando la distribución de la censura (Exponencial, Uniforme), con porcentajes de censura 50 %, 43 %, 23 % y 22 %, compara estos intervalos de varianza estabilizada, Brokmeyer-Crowley, reflexivos simples y reflexivos transformados. Con $n = 21$, también intervalos de confianza repetidos, con distribución de tiempos de supervivencia (Exponencial y Weibull), utilizando 5000 simulaciones y porcentajes de censura de 5 % y 10 %, presenta una aplicación con porcentajes de censura 56 %, 68 %, 83 %, 66 % y 53 %, y compara los intervalos de confianza Pocock, Obrien y Fleming, usando tamaños de muestra fija.

Slud, Byar y Green (1984), comparan intervalos de confianza reflexivos Efron, Emerson, Reflexivos transformados, con Brookmeyer-Crowley, Simon y Lee, Unsmoothed, Reid, usando tamaños de muestra $n = 21, 22, 25, 40, 41, 42, 60, 61, 62$ y $\alpha = 0.05$ y 0.1 , también utilizan la distribución de los tiempos de supervivencia exponenciales y Weibull, y para la censura Exponenciales y Uniformes, con porcentajes de censura esperados de (50,43.2,23.1,22,47.6,42.3,24.6,24.9,52.7,43.8,22.5,20).

Anderson, Bernstein y Pike (1982), comparan intervalos de confianza para la función de supervivencia mediante la formula de Greenwood, Rothman-Wilson, transformaciones *logit*, *Arcsen* y *log(-log)*, utilizando distribución de los tiempos/censura Weibull, con tamaños de muestra de $n = 25, 50, 100$ y 200 y porcentajes de censura de 0% , 25% y 50% .

Brookmeyer y Crowley (1982), desarrollan intervalos de confianza noparamétricos para la función de supervivencia, mediante la generalización de la prueba del signo, comparan intervalos de confianza (noparamétricos, Bartolomew y de varianza estabilizada), utilizando tiempos de supervivencia distribuidos exponencial y Weibull con porcentajes de censura de 15% , 20% , 37% y 52% y nivel $\alpha = 0.01, 0.05, 0.1, 0.2$ y 0.25 , muestra la longitud del intervalo, también utiliza la distribución Reyleigh con porcentajes de censura de 14% , 18% y 36% , por otra parte compara intervalos de confianza para 4 tratamientos con tamaños de muestra (56, 58, 52) y porcentajes de censura de 7% , 8% , 14% y 16% , mediante una formula empírica.

Emerson (1982), propone un método para construir intervalos de confianza noparamétricos para la mediana de la distribución de supervivencia, utilizando el test del signo, compara los intervalos de confianza noparamétricos y los de varianza estabilizada usando porcentajes de censura 12% , 20% , 37% y 52% , utilizando distribuciones para los tiempos de supervivencia Exponencial y Weibull, a través de la cobertura y la longitud del intervalo resultante, para tamaños de muestra $n = 25$ y 50 con 1600 réplicas.

Reid (1981), propone dos métodos para encontrar intervalos de confianza para el tiempo mediano, usando métodos Bootstrap, con $n = 97$ y 46 observaciones censuras, en algunos puntos percentiles, mediante la ecuación condicional e incondicional, aproximación normal usando replicas de $100, 200, 300$ y 400 y porcentajes de censura de 71% , 70% y 49% y tamaños de muestra de $21, 23$ y 51 , presentando la longitud del intervalo y la confianza.

Tomas y Grunkemeier (1975), calculan intervalos de confianza para las probabilidades de supervivencia con $n = 25, 50, 100$ y 200 , con porcentajes de censura 5% , 10% , 25% y 50% utilizando la distribución de los tiempos de supervivencia Weibull.

1.1. Funciones de Básicas

Sea T una variable aleatoria no negativa que representa el tiempo de vida de los individuos de alguna población. Usualmente, se asume que T es continua, en el intervalo $[0, \infty)$. La distribución de probabilidad para un tiempo de falla T puede ser caracterizado por una función de distribución acumulativa, una función de densidad de probabilidad, una función de supervivencia y una función hazard.

La escogencia de la función o funciones a utilizar, depende de la conveniencia para el modelo de especificación, la interpretación o el desarrollo técnico. Todas estas funciones son importantes de una u otra manera, Lawless, J. (2003). Esta variable genérica puede ser expresada en diferentes unidades, tales como, uso en ciclos, horas, años, etc).

Función de distribución acumulativa

La función de distribución acumulativa de T (fda) se define como:

$$F(t) = P(T \leq t) \quad (1.1)$$

Es la probabilidad de que una unidad falle antes del tiempo t . Esta probabilidad puede ser interpretada como la proporción de unidades en una población que fallarán antes del tiempo t .

Función de supervivencia

La función de supervivencia también se llama función de confiabilidad, y permite obtener la probabilidad de supervivencia mas allá del tiempo t . Esta función se define como:

$$S(t) = P(T > t) \quad (1.2)$$

Es el complemento de la fda. Por lo tanto $F(t) + S(t) = 1$, para todo t . Donde $F(t)$ y $S(t)$ son funciones creciente y decreciente respectivamente.

Función de densidad de probabilidad

La función de densidad de probabilidad (fdp) para la variable aleatoria continua T se define como la derivada de $F(t)$ con respecto a t :

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \quad (1.3)$$

La fdp puede ser utilizada para representar la frecuencia relativa de tiempos de falla como una función del tiempo.

Función Hazard

Los estudios modernos de confiabilidad están basados en las funciones hazard. La función hazard o tasa hazard es una función definida como

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (1.4)$$

Esta función de riesgo, especifica la tasa instantánea de muerte o falla en un intervalo de tiempo, dado que el individuo sobrevivió hasta t . En particular, $h(t)\Delta t$ es la probabilidad aproximada de muerte en $[t, t + \Delta t)$, dado que el sujeto ha sobrevivido hasta t .

La función de riesgo es denominada también, razón de riesgo, razón de falla, y fuerza de mortalidad.

Función Hazard Acumulativa

La función hazard acumulativa se define como:

$$H(t) = \int_0^t h(s) ds \quad (1.5)$$

donde

$$h(t) = \frac{-d(\log S(t))}{dt} \quad (1.6)$$

entonces,

$$S(t) = \exp\left(-\int_0^t h(s)ds\right) = \exp(-H(t)) \quad (1.7)$$

1.1.1. Censura

Un estudio del tiempo de supervivencia por lo general se inicia mediante la recopilación de observaciones en el tiempo, donde los datos no se recogen en un solo día o en la mayoría de los casos, ni siquiera durante un corto período de tiempo. Más bien, como avanza el estudio, cada sujeto se identifica a medida que entra en el estudio. Los individuos observados o bien están vivos al final del estudio o mueren en algún momento durante el estudio, Selvin (2008). Entonces la censura es un fenómeno que se presenta a menudo cuando se estudia el tiempo de supervivencia y representa una información incompleta del tiempo de supervivencia exacto, así, estos datos censurados representan información parcial sobre la variable a estudiar.

1.1.2. Tipos de Censura

Censura Aleatoria:

Corresponde cuando algún paciente muere por causas ajenas al evento de interés en el experimento, por ejemplo se observa a un grupo de pacientes con un nuevo tratamiento que mejora su supervivencia a cierta enfermedad. Un paciente muere en un accidente de tráfico por lo que es una observación con censura aleatoria.

Censura de Intervalo:

Este tipo de censura refleja la incertidumbre al momento exacto de ocurrencia del evento en las unidades dentro de un intervalo de tiempo. Este tipo de datos aparece con frecuencia de las pruebas o situaciones en las que los objetos de interés no son de seguimiento constante. Si se está ejecutando una prueba en ciertas unidades y su

inspección es cada t horas, sólo se sabe que una unidad falló o no entre las inspecciones.

Censura a Derecha:

Denominada como censura de tipo I, consiste cuando cada individuo tiene un tiempo potencial de censura $C_i > 0$ tal que T_i es observado si $T_i \leq C_i$, de lo contrario, sabemos sólo que $T_i > C_i$. La censura Tipo I surge a menudo cuando se lleva a cabo un estudio durante un período de tiempo especificado, Lawless (2003).

También se dice que una observación está censurada por la derecha en t_c cuando sólo se conoce si su valor es mayor o igual que t_c pero no se sabe su valor exacto.

Como notación tenemos que

$$t_i = \min(T_i, C_i), \quad \delta_i = I(T_i \leq C_i)$$

Para la censura Tipo I, la función de verosimilitud para una muestra censurada está basada en la distribución de probabilidad de (t_i, δ_i) , $i = 1, 2, \dots, n$. Ambas t_i y δ_i son variables aleatorias y su función de densidad de probabilidad conjunta es

$$f(t_i)^{\delta_i} P(T_i > C_i) \tag{1.8}$$

Notemos que C_i son constantes fijas y que t_i pueden tomar valores menores ó iguales a C_i con

$$P(t_i = C_i, \delta_i = 0) = P(T_i > C_i)$$

$$P(t_i, \delta_i = 1) = f(t_i), \quad t_i \leq C_i$$

Donde $P(\cdot)$ en la segunda expresión denota una p.d.f cualquiera o función de masa de probabilidad según si la distribución de T_i es continua o discreta sobre t_i . Asumiendo que los tiempos de vida T_1, \dots, T_n son estadísticamente independientes, obtenemos la función de verosimilitud para (1.8) como

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i+)^{1-\delta_i} \tag{1.9}$$

El término $S(t_i+)$ corresponde a la $P(T_i > t_i)$, en general; si $S(t)$ es continua sobre t_i , entonces $S(t_i+) = S(t_i)$.

1.2. Intervalos de Confianza paramétricos

1.2.1. Caso Exponencial:

Para el caso exponencial, supongamos que los tiempos de vida T_i son independientes y siguen una distribución exponencial con función de densidad de probabilidad $f(t) = \lambda \exp(-\lambda t)$ y función de supervivencia $S(t) = \exp(-\lambda t)$, entonces (1.9) da como resultado

$$L(\lambda) = \prod_{i=1}^n (\lambda e^{-\lambda t_i})^{\delta_i} (e^{-\lambda t_i})^{1-\delta_i} = \lambda^r \exp(-\lambda \sum_{i=1}^n t_i) \quad (1.10)$$

Donde $r = \sum \delta_i$ es el número de tiempos de vida no censurados. La función de log-verosimilitud $l(\lambda) = \log(L(\lambda))$ es

$$l(\lambda) = r \log(\lambda) - \lambda \sum_{i=1}^n t_i \quad (1.11)$$

La estimación de máxima verosimilitud es dada por la solución $\frac{dl}{d\lambda} = 0$, y corresponde a

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^n t_i} \quad (1.12)$$

para encontrar un intervalo de confianza $100(1 - \alpha)\%$ para un tiempo de interés t se debe hallar:

$$\text{Var}(S(t)) = (\lambda t)^2 \exp(-2\lambda t) \left(\frac{1}{\lambda^2} \right) = t^2 \exp(-2\lambda t) = t^2 S^2(t) \quad (1.13)$$

entonces el I.C esta dado por

$$\left[S(t) - Z_{(1-\frac{\alpha}{2})} t S(t), S(t) + Z_{(1-\frac{\alpha}{2})} t S(t) \right] \quad (1.14)$$

por lo tanto los límites de confianza son:

$$S(t) \left[1 \pm Z_{(1-\frac{\alpha}{2})} t \right] \quad (1.15)$$

Sin pérdida de generalidad digamos que para un tiempo de interés como la mediana $t_{0,5} = M$, con base en los resultados de Bartholomew (1957) puede ser estimada por

$$\hat{M} = \frac{\ln(2) \sum(\text{tiempos supervivencia observados})}{d} \quad (1.16)$$

en el caso de tiempos de vida exponencial, donde d corresponde al numero total de fallas.

1.2.2. Caso Weibull:

En el caso Weibull, suponemos que los tiempos de vida T_i son independientes y siguen una distribución Weibull con función de densidad de probabilidad $f(t) = \lambda\beta(\lambda t)^{\beta-1} \exp[-(\lambda t)^\beta]$ y función de supervivencia $S(t) = \exp[-(\lambda t)^\beta]$. Entonces (1.9) da como resultado

$$(\lambda\beta)^r \prod_{i=1}^r (\lambda t_i)^{\beta-1} \exp \left\{ - \left(\frac{1}{\lambda} \right)^{-\beta} \sum_{i=1}^r t_i^\beta - \left(\frac{1}{\lambda} \right)^{-\beta} (n-r)t_c^\beta \right\} \quad (1.17)$$

luego el logaritmo de la verosimilitud es

$$r \log \beta - r \log \left(\frac{1}{\lambda} \right) + (\beta-1) \sum_{i=1}^r [\log t_i - \log \left(\frac{1}{\lambda} \right)] - \left(\frac{1}{\lambda} \right)^{-\beta} \sum_{i=1}^r t_i^\beta - \left(\frac{1}{\lambda} \right)^{-\beta} (n-r)t_c^\beta \quad (1.18)$$

entonces los estimadores de máxima verosimilitud se obtienen resolviendo el sistema de las dos ecuaciones siguientes

$$\left[\frac{\sum_{i=1}^r t_i^\beta \log(t_i) + (n-r)t_c^\beta \log(t_c)}{\sum_{i=1}^r t_i^\beta + (n-r)t_c^\beta} - \frac{1}{\beta} \right] - \frac{1}{r} \sum_{i=1}^r \log(t_i) = 0 \quad (1.19)$$

y

$$\left[\frac{\sum_{i=1}^r t_i^\beta + (n-r)t_c^\beta}{r} \right] = \left(\frac{1}{\lambda} \right)^\beta \quad (1.20)$$

donde

$$Var(t) = \left(\frac{1}{\lambda^2}\right) \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right]$$

Además

$$Var(S(t)) = Var\left(\exp[-(\lambda t)^\beta]\right) = \left[\beta(\lambda t)^{\beta-1}\right]^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right] \quad (1.21)$$

Entonces el I.C $100(1 - \alpha)\%$ para $S(t)$ es:

$$\left[S(t) - Z_{(1-\frac{\alpha}{2})} \sqrt{Var(S(t))}, S(t) + Z_{(1-\frac{\alpha}{2})} \sqrt{Var(S(t))} \right] \quad (1.22)$$

así, los límites están dados por

$$S(t) \pm Z_{(1-\frac{\alpha}{2})} \left[\beta(\lambda t)^{\beta-1} \right] \sqrt{\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right]} \quad (1.23)$$

1.3. Estimadores no paramétricos

1.3.1. Estimador de Kaplan Meier Para la Función De Supervivencia

Una manera útil de retratar los datos de supervivencia es calcular y graficar la función de supervivencia empírica o equivalentemente, la función de distribución empírica. Esto también proporciona una estimación no paramétrica de la supervivencia o función de distribución para la distribución de los tiempos de vida bajo estudio. Si no existe observaciones censuradas en una muestra de tamaño n , la función de supervivencia empírica (FSE) es definida como:

$$\hat{S}(t) = \frac{\# \text{ de observaciones } \geq t}{n} \quad \text{con } t \geq 0 \quad (1.24)$$

Esto es una función escalonada que decrece en $\frac{1}{n}$ justo después de cada observación del tiempo de vida si todas las observaciones son distintas. Más generalmente, si hay

d tiempos de vida igual a t , la FSE decrece $\frac{d}{n}$ justamente pasado t , Lawless (2003)

Cuando se trata con datos censurados, es necesario algunas modificaciones en la anterior ecuación, puesto que el número de tiempos de vida mayor o igual a t generalmente no se conocerá exactamente. Kaplan y Meier (1958), propusieron una modificación de $\hat{S}(t)$ a la cual denominaron, estimador del producto límite (EPL) de la función de supervivencia.

Supónganse que existen observaciones de n individuos y que hay $k \leq n$ tiempos distintos en los cuales la muerte ocurre, esto es, $t_1 < t_2 < \dots < t_k$.

Se admite la posibilidad de tener más de una muerte en t_j y d_j representará el número de muertes en t_j . Además, existen los tiempos de censura t_c para individuos cuyo tiempo de vida no es observado.

El estimador del producto límite, $\hat{S}(t)$, es definido como:

$$\hat{S}(t) = \prod_{j:t_j < t} \frac{n_j - d_j}{n_j} \quad (1.25)$$

Donde n_j es el número de individuos en riesgo en t_j , es decir, el número de individuos vivos y no censurados justo antes de t_j . Cualquier individuo con tiempo de censura registrado igual a t_j será incluido en el conjunto de n_j individuos en riesgo en t_j , como individuos que murieron en t_j . Esta convención es razonable, puesto que un individuo censurado en el tiempo t_c casi ciertamente sobrevive después de t_c .

Propiedades del Estimador del Producto Límite

El estimador del producto límite $\hat{S}(t)$ posee un número importante de propiedades, la principal es que $\hat{S}(t)$ es un estimador consistente de $S(t)$ bajo condiciones realmente generales; $n \rightarrow \infty$. En la definición de $\hat{S}(t)$ se asume que existen k tiempos de vida distintos $t_1 < t_2 < \dots < t_k$ con d_j muertes y n_j individuos en riesgo en t_j . Hay además algunos tiempos censurados.

Supongamos que en el intervalo $[t_{j-1}, t_j)$ hay observados λ_j tiempos censurados. Haciendo $t_0 = 0$, $t_{k+1} = \infty$, $j = 1, 2, \dots, k + 1$, se incluye tiempos de censura que ocurren antes de la primera o después de la última observación de tiempo de vida.

La función de supervivencia teórica para tiempos de vida es $S(t)$, y la probabilidad de

morir de un individuo en t_j es $S(t_j) - S(t_j + 0)$, donde $S(t)$ es una función no creciente continúa por la izquierda. Bajo las suposiciones con respecto a los tipos de censuras¹, la función de verosimilitud observada es de la forma.

$$L = \prod_{d \in D} [S(t_j) - S(t_j + 0)] \prod_{r \in R} [S(t_r)] \quad (1.26)$$

donde D y R son los conjuntos de tiempo de falla exacto y tiempos de falla censurados a derecha, respectivamente.

Estimación de la Varianza Kaplan-Meier

Para evaluar los resultados eficazmente cuando se usa el estimador del producto límite (1.25) es conveniente tener un estimador de la varianza de $\hat{S}(t)$, una de las utilizadas en este trabajo es la formula de Greenwood (1926), la cual está dada por:

$$Var(\hat{S}(t)) = \hat{S}^2(t) \sum_{j:t_j < t} \frac{d_j}{n_j[n_j - d_j]} \quad (1.27)$$

La derivación de (1.27) es:

$$\log(\hat{S}(t)) = \log \left(\prod_{j:t_j < t} \frac{n_j - d_j}{n_j} \right) = \sum_{j=1}^k \log \left(\frac{n_j - d_j}{n_j} \right) = \sum_{j=1}^k \log(p_j) \quad (1.28)$$

Entonces

$$Var(\log(\hat{S}(t))) = \sum_{j=1}^k Var(\log(p_j)) \quad (1.29)$$

Aplicando el método delta

$$Var(\log(p_j)) \approx \left(\frac{1}{\pi_j} \right)^2 \frac{\pi_j(1 - \pi_j)}{n_j} = \left(\frac{1}{\pi_j} \right) \frac{1 - \pi_j}{n_j} \quad (1.30)$$

Entonces

$$Var(\log(\hat{S}(t))) \approx \sum_{j=1}^k \left(\frac{1}{\pi_j} \right) \frac{1 - \pi_j}{n_j} \quad (1.31)$$

¹censura por la derecha(Tipo I)

Ahora

$$Var(\hat{S}(t)) \approx [S(t)]^2 Var(\log(\hat{S}(t))) = [S(t)]^2 \sum_{j=1}^k \left(\frac{1}{\pi_j} \right) \frac{1 - \pi_j}{n_j} \quad (1.32)$$

Finalmente la formula de Greenwood, se obtiene reemplazando $p_j = \frac{n_j - d_j}{n_j}$ por π_j .

Por lo tanto

$$Var(\hat{S}(t)) = \hat{S}^2(t) \sum_{j:t_j < t} \frac{d_j}{n_j[n_j - d_j]}$$

1.3.2. Estimador de Nelson-Aalen Para la Función De Supervivencia

Nelson y Aalen (1969,1972), sugieren otra alternativa para estimar la función de supervivencia basándose en la estimación de la función Hazard Acumulada como

$$\hat{H}(t) = \sum_{j:t_j < t} \frac{d_j}{n_j} \quad (1.33)$$

este estimador es muy utilizado en los casos cuando se tienen tamaños de muestra pequeños, donde

$$\hat{S}(t) = \exp(-\hat{H}(t))$$

Estimación de la Varianza Estimador Nelson-Aalen

Una forma de calcular la varianza la función de supervivencia, basándose en la función Hazard Acumulada es

$$\widehat{Var}(\hat{H}(t)) = \sum_{j:t_j < t} \frac{d_j(n_j - d_j)}{n_j^3} \quad (1.34)$$

otra forma alternativa para varianza es

$$\widehat{Var}(\hat{H}(t)) = \sum_{j:t_j < t} \frac{d_j}{n_j^2} \quad (1.35)$$

Aalen y Johansen (1978), proponen una alternativa para la estimación de la varianza de $\hat{S}(t)$, como

$$\widehat{Var}(\hat{S}(t)) = \hat{S}^2(t) \sum_{j:t_j < t} \frac{d_j}{n_j^2} \quad (1.36)$$

La cual es utilizada en éste trabajo.

1.4. Intervalos de confianza no paramétricos

1.4.1. Intervalo de Confianza mediante el estimador de Kaplan Meier

Teniendo en cuenta la normalidad asintótica de los estimadores de máxima verosimilitud bajo condiciones de regularidad, los intervalos de $100(1 - \alpha)\%$ de confianza de la función de supervivencia en cada tiempo fijo t_i pueden ser calculados de la siguiente manera

$$\left(\hat{S}(t_j) - Z_{(1-\frac{\alpha}{2})} \sqrt{Var[\hat{S}(t_j)]} < S(t_j) < \hat{S}(t_j) + Z_{(1-\frac{\alpha}{2})} \sqrt{Var[\hat{S}(t_j)]} \right) \quad (1.37)$$

donde $Z_{(1-\alpha/2)}$ es el valor que se excede con probabilidad $(1-\alpha/2)$ para una distribución normal estándar.

Luego para el caso del tiempo de interés $t_i = 0.2$, el IC $100(1 - \alpha)\%$ es

$$\left(\hat{S}(0.2) - Z_{(1-\frac{\alpha}{2})} \sqrt{Var[\hat{S}(0.2)]}, \hat{S}(0.2) + Z_{(1-\frac{\alpha}{2})} \sqrt{Var[\hat{S}(0.2)]} \right) \quad (1.38)$$

donde

$$\widehat{Var}[\hat{S}(0.2)] = \hat{S}^2(0.2) \sum_{j:t_j < 0.2} \frac{d_j}{n_j[n_j - d_j]} \quad (1.39)$$

1.4.2. Intervalo de Confianza mediante el estimador de Nelson-Aalen

Un intervalo de confianza $100(1-\alpha)\%$ para 0.2, mediante el estimador de Nelson-Aalen esta dado por

$$\left(\hat{S}(0.2) - Z_{(1-\frac{\alpha}{2})} \sqrt{Var[\hat{S}(0.2)]}, \hat{S}(0.2) + Z_{(1-\frac{\alpha}{2})} \sqrt{Var[\hat{S}(0.2)]} \right) \quad (1.40)$$

donde

$$\widehat{Var}(\hat{S}(0.2)) = \hat{S}^2(0.2) \sum_{j:t_j < 0.2} \frac{d_j}{n_j^2} \quad (1.41)$$

1.4.3. Intervalos de Confianza Mediante Transformaciones

En algunos casos es de mucho interés calcular intervalos de confianza, usando transformaciones *log*, *log(-log)*, *logit* y *Arco seno*, con el hecho principal ya sea de simetrizar la distribución de un parámetro cualquiera θ , estabilizar la varianza, etc.

IC Mediante transformación *log*

Estos intervalos de confianza fueron sugeridos inicialmente por Kalbfleish y Prentice (1980), luego un intervalo de confianza $100(1-\alpha)\%$ mediante esta transformación para $S(0.2)$ es

$$\left(\hat{S}(0.2) \exp \left\{ \frac{-Z_{(1-\frac{\alpha}{2})} \sigma_s(0.2)}{\hat{S}(0.2)} \right\}, \hat{S}(0.2) \exp \left\{ \frac{Z_{(1-\frac{\alpha}{2})} \sigma_s(0.2)}{\hat{S}(0.2)} \right\} \right) \quad (1.42)$$

donde

$$\sigma_s^2(0.2) = \frac{\widehat{Var}[\hat{S}(0.2)]}{\hat{S}^2(0.2)} \quad (1.43)$$

es decir que $\sigma_s(0.2)$ corresponde a la raíz de la suma en la formula de Greenwood (1.39).

IC Mediante transformación $\log(-\log)$

Para encontrar intervalos de confianza $100(1 - \alpha)\%$ mediante esta transformación, se deben de encontrar los límites

$$\hat{S}(0.2)^{\exp\left\{\pm Z_{(1-\frac{\alpha}{2})} \frac{\sigma_s(0.2)}{\log(\hat{S}(0.2))}\right\}} \quad (1.44)$$

tales que

$$\left(\hat{S}(0.2)^{\exp\left\{-Z_{(1-\frac{\alpha}{2})} \frac{\sigma_s(0.2)}{\log(\hat{S}(0.2))}\right\}}, \hat{S}(0.2)^{\exp\left\{Z_{(1-\frac{\alpha}{2})} \frac{\sigma_s(0.2)}{\log(\hat{S}(0.2))}\right\}} \right) \quad (1.45)$$

IC Mediante transformación *Arco seno*

Otra alternativa para calcular intervalo de confianza $100(1 - \alpha)\%$ para la función de supervivencia es mediante la transformación *Arcsen*, sugeridos inicialmente por Nair (1984), están dados por

$$(LI < S(0.2) < LS)$$

donde

$$LI = \text{sen}^2 \left\{ \max \left\langle 0, \arcsen(\sqrt{\hat{S}(0.2)}) - 0,5Z_{(1-\frac{\alpha}{2})}\sigma_s(0.2)\sqrt{\frac{\hat{S}(0.2)}{1-\hat{S}(0.2)}} \right\rangle \right\} \quad (1.46)$$

y

$$LS = \text{sen}^2 \left\{ \min \left\langle \frac{\pi}{2}, \arcsen(\sqrt{\hat{S}(0.2)}) + 0,5Z_{(1-\frac{\alpha}{2})}\sigma_s(0.2)\sqrt{\frac{\hat{S}(0.2)}{1-\hat{S}(0.2)}} \right\rangle \right\} \quad (1.47)$$

1.4.4. El Método de Remuestreo Bootstrap Aplicado a datos de Supervivencia

El *Bootstrap* ha sido propuesto como un método de análisis en varias situaciones relacionadas a datos de supervivencia, particularmente donde hay complicaciones como

datos censurados. Por ejemplo, un estudio médico puede involucrar registros del tiempo de supervivencia de pacientes que tuvieron una operación de alto riesgo.

Un problema involucra determinar los límites de confianza para la función de supervivencia teórica o parámetros que describen esta función. Efron (1986), sugiere utilizar el método bootstrap para estimar la supervivencia utilizando datos de registros diarios de muertes y el estimador de Kaplan Meier.

El desarrollo y discusión se proporcionan por Efron (1986), y Akritas (1986). En particular Akritas (1986) y Efron (1981) discutieron la utilidad del método del remuestreo bootstrap en la construcción de bandas de confianza para una función de supervivencia, con un cierto nivel de confianza.

Para estimar los intervalos de confianza para Función de Supervivencia (1.2), utilizando el estimado de Kaplan-Meier en (1.25), a través del intervalo de confianza (1.38), mediante el remuestreo bootstrap, consiste en lo siguiente:

1. Dada la muestra de tamaño n , estimar $\hat{S}(0.2)$. La distribución de esta muestra se considera equivalente a la distribución de la población y $\hat{S}(0.2)$ es el estimador muestral del parámetro poblacional $S(0.2)$.
2. Generar B muestras *bootstrap* de tamaño n mediante muestreo con reemplazamiento de la muestra original, asignando a cada tiempo una probabilidad $1/n$ y calcular los correspondientes valores $\hat{S}(0.2)^{*1}, \hat{S}(0.2)^{*2}, \dots, \hat{S}(0.2)^{*B}$ para cada una de las B muestras bootstrap.
3. Estimar el error estándar del parámetro estimado $\hat{S}(0.2)$ calculando la desviación estándar de las B réplicas bootstrap.

Así, obtenemos que el error estándar es

$$\sigma_{S(0.2)}^* = \sqrt{\frac{\sum_{b=1}^B (S(0.2)^{b*} - \bar{\hat{S}}(0.2)^*)^2}{(B-1)}} = \sigma_{BOOT} \quad (1.48)$$

donde

$$\bar{\hat{S}}(0.2)^* = \frac{1}{B} \sum_{b=1}^B S(0.2)^{b*} \quad (1.49)$$

Luego el intervalo de confianza $100(1 - \alpha) \%$ para $S(0.2)$ esta dado por

$$\left(\hat{S}(0.2) - Z_{\frac{\alpha}{2}} \sigma_{BOOT}, \hat{S}(0.2) + Z_{\frac{\alpha}{2}} \sigma_{BOOT} \right) \quad (1.50)$$

donde σ_{BOOT} esta dado en (1.48) y $\hat{S}(0.2)$ corresponde al estimador Kaplan-Meier. De forma análoga usando el estimador de Nelson-Aalen.

1.4.5. Criterio de Comparación de los Intervalos de Confianza

Hay dos conceptos importantes que se deben de considerar al evaluar los intervalos de confianza: la precisión indicada por la longitud del intervalo y la probabilidad de cobertura $P(L_{inf} \leq S(t) \leq L_{sup})$. Estos criterios no se pueden analizar por separado porque de poco nos sirve un intervalo con probabilidad de cobertura alta si su longitud es muy grande o un intervalo con una longitud muy pequeña pero con probabilidad de cobertura muy baja.

Idealmente se quiere ver que los intervalos sean cortos y tengan probabilidad de cobertura muy cercana al nivel de confianza nominal, que los procedimientos que se utilicen para construir los intervalos de confianza den intervalos tales que: sus longitudes sean pequeñas, pero diferentes de cero y que la probabilidad de cobertura no sea inferior al nivel de confianza nominal, Correa y Sierra (2003).

En este trabajo se calcula para cada método las tasas de error (TE), la longitud promedio del intervalo de confianza (LPI) para la función de supervivencia y la propuesta de Correa et al., (2003), del índice de comparación de intervalos de confianza el cual tiene en cuenta simultáneamente el nivel de confianza real (NR), el nivel de confianza nominal (NN), la longitud promedio del intervalo (LPI) como

$$I = \frac{2 - LPI}{2} \times \frac{NR}{NN} \quad (1.51)$$

donde el nivel de confianza real(NR): corresponde a la proporción de intervalos simulados que cubre el verdadero valor de $S(t)$. Por lo tanto mientras mayor sea el índice mejor será el método, luego:

$$T.E = \frac{\# \text{ de I.C que no cubren el verdadero valor } S(t)}{N}, \quad NR = 1 - T.E \quad \text{y} \quad LPI = \sum_{i=1}^N \frac{(LS_i - LI_i)}{N}$$

donde, N: corresponde al número de simulaciones, LS y LI representa el límite superior e inferior respectivamente.

Capítulo 2

ESTUDIO DE SIMULACIÓN

En este capítulo se presentan los resultados obtenidos al comparar los intervalos de confianza para la función de supervivencia en un tiempo de interés $t_i = 0.2$, este tiempo de interés fue escogido teniendo en cuenta la literatura y con el propósito de ser comparables con los resultados Borgan y Liestol (1990), mediante los estimadores de Kaplan-Meier y Nelson-Aalen, a través de los métodos de remuestreo *Bootstrap* aplicado a los estimadores, transformaciones *log*, *log(-log)* y *Arco seno*. Estos métodos se compararon utilizando un algoritmo en R descrito el capítulo 5.

2.1. Escenarios de Simulación

En esta sección se describen los escenarios de simulación utilizados para la comparación de los intervalos de confianza para función de supervivencia en un tiempo de interés, se utilizaron diferentes combinaciones de modelos generadores de los tiempos de supervivencia y de censura, como son; modelo Exponencial(α) y Weibull(α, β), donde $\alpha = 2, 1$ y $\beta = 0.5$ con porcentajes de censura tipo I, de 0 %, 15 %, 25 %, 35 %, 45 %, 55 % y tamaños de muestra $n = 25, 50, 75$ y 100.

La razón de utilizar los porcentajes de censura y los tamaños de muestra se debe a la finalidad de comparar los resultados con los resultados de algunos autores referenciados en el capítulo 2. Además con el propósito de abreviar los títulos de las tablas se presenta las siguientes siglas, TE: Tasas de Error, LPI: Longitud promedio del intervalo y NN: Nivel de confianza nominal, LI.PAR: representa la longitud del I.C parametrico (real), T.KM: I.C mediante el estimador de Kaplan-Meier, Boot.KM y Boot.NA: I.C mediante

el remuestreo *Bootstrap* a los estimadores de Kaplan-Meier y Nelson-Aalen respectivamente, además $\log \log(-\log)$ y *Arco seno* corresponden a las transformaciones de las funciones de supervivencia definidas en (1.42), (1.45) y (1.46) respectivamente. Entonces para determinar la efectividad de utilizar intervalos de confianza para la función de supervivencia no paramétricos, frente a los paramétricos, los valores de referencia para el caso exponencial $exp(2)$ la $S(t) = 0.67032$ y $exp(1)$ la $S(t) = 0.8187308$, mientras que para el caso $weib(1, 0.5)$ la $S(t) = 0.6394073$ se presentan a continuación los resultados:

2.2. Resultados

2.2.1. Resultados para $n = 25$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.053(0.37)	0.080(0.39)	0.017(0.63)	0.076(0.37)	0.053(0.36)
Exp/Weib	(0.526)	0.029(0.32)	0.090(0.32)	0.005(0.42)	0.070(0.33)	0.074(0.31)
Weib/Exp	(0.315)	0.102(0.37)	0.080(0.38)	0.022(0.64)	0.047(0.37)	0.061(0.36)
Weib/Weib	(0.315)	0.102(0.37)	0.080(0.38)	0.022(0.64)	0.047(0.37)	0.061(0.36)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.053(0.36)	0.110(0.36)	0.017(0.62)	0.076(0.37)	0.053(0.36)
Exp/Weib	(0.526)	0.029(0.31)	0.100(0.30)	0.005(0.41)	0.070(0.33)	0.074(0.30)
Weib/Exp	(0.315)	0.077(0.37)	0.100(0.36)	0.022(0.62)	0.047(0.37)	0.036(0.36)
Weib/Weib	(0.315)	0.077(0.37)	0.100(0.36)	0.022(0.62)	0.047(0.37)	0.036(0.36)

Tabla 2.1: TE y LPI de NN 95 % con 0 % de censura para $n = 25$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.069(0.36)	0.090(0.38)	0.056(0.56)	0.033(0.36)	0.053(0.35)
Exp/Weib	(0.526)	0.040(0.30)	0.090(0.32)	0.027(0.38)	0.048(0.32)	0.046(0.29)
Weib/Exp	(0.315)	0.113(0.36)	0.150(0.37)	0.063(0.56)	0.028(0.36)	0.067(0.35)
Weib/Weib	(0.315)	0.107(0.36)	0.130(0.38)	0.071(0.58)	0.034(0.37)	0.076(0.36)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.066(0.35)	0.110(0.35)	0.057(0.54)	0.031(0.37)	0.061(0.35)
Exp/Weib	(0.526)	0.038(0.30)	0.100(0.30)	0.027(0.38)	0.034(0.32)	0.046(0.29)
Weib/Exp	(0.315)	0.128(0.36)	0.160(0.35)	0.063(0.55)	0.027(0.37)	0.065(0.35)
Weib/Weib	(0.315)	0.122(0.36)	0.130(0.36)	0.071(0.56)	0.031(0.37)	0.075(0.35)

Tabla 2.2: TE y LPI de NN 95 % con 15 % de censura para $n = 25$

Notamos en el primer escenario que cuando no hay observaciones censuradas los intervalos de confianza mediante la transformación $\log(S(t))$ presentan mejor comportamiento, esto se debe a que resultan ser más amplios, independientemente del estimador que se utilice y el modelo generador.

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.112(0.35)	0.160(0.37)	0.010(0.52)	0.024(0.36)	0.075(0.34)
Exp/Weib	(0.526)	0.058(0.29)	0.136(0.31)	0.051(0.36)	0.035(0.31)	0.053(0.29)
Weib/Exp	(0.315)	0.165(0.35)	0.112(0.32)	0.114(0.53)	0.041(0.36)	0.114(0.34)
Weib/Weib	(0.315)	0.145(0.36)	0.131(0.33)	0.124(0.54)	0.047(0.36)	0.119(0.34)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.113(0.35)	0.172(0.34)	0.106(0.51)	0.021(0.36)	0.089(0.34)
Exp/Weib	(0.526)	0.057(0.29)	0.152(0.29)	0.051(0.35)	0.024(0.32)	0.054(0.28)
Weib/Exp	(0.315)	0.189(0.35)	0.161(0.33)	0.115(0.51)	0.040(0.36)	0.114(0.34)
Weib/Weib	(0.315)	0.165(0.35)	0.165(0.34)	0.125(0.53)	0.045(0.37)	0.124(0.34)

Tabla 2.3: TE y LPI de NN 95 % con 25 % de censura para $n = 25$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.202(0.34)	0.254(0.35)	0.172(0.47)	0.045(0.35)	0.134(0.33)
Exp/Weib	(0.526)	0.118(0.27)	0.142(0.30)	0.116(0.33)	0.027(0.31)	0.096(0.27)
Weib/Exp	(0.315)	0.266(0.34)	0.250(0.36)	0.212(0.47)	0.091(0.35)	0.210(0.33)
Weib/Weib	(0.315)	0.253(0.35)	0.231(0.35)	0.227(0.50)	0.091(0.36)	0.208(0.34)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.208(0.33)	0.282(0.33)	0.191(0.46)	0.045(0.35)	0.153(0.32)
Exp/Weib	(0.526)	0.118(0.27)	0.178(0.28)	0.116(0.32)	0.025(0.31)	0.101(0.27)
Weib/Exp	(0.315)	0.292(0.33)	0.243(0.35)	0.2135(0.46)	0.091(0.35)	0.211(0.33)
Weib/Weib	(0.315)	0.273(0.34)	0.252(0.35)	0.234(0.48)	0.091(0.36)	0.223(0.33)

Tabla 2.4: TE y LPI de NN 95 % con 35 % de censura para $n = 25$

entonces a medida que se aumenta el porcentaje de censura los intervalos de confianza para la función de supervivencia mediante la transformación $\log(-\log(S(t)))$ resultan ser más efectivos. Por otra parte se puede afirmar que en general para muestras de tamaño $n = 25$ a medida que se aumentan las observaciones censuradas los intervalos de confianza para la función de supervivencia en el tiempo de interés $t = 0.2$ utilizando el estimador de Kaplan-Meier las tasas de error y las longitudes de los intervalos son mas pequeñas.

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.313(0.33)	0.330(0.34)	0.262(0.43)	0.075(0.34)	0.205(0.32)
Exp/Weib	(0.526)	0.172(0.26)	0.170(0.29)	0.171(0.31)	0.040(0.30)	0.136(0.26)
Weib/Exp	(0.315)	0.390(0.33)	0.410(0.35)	0.342(0.44)	0.153(0.34)	0.330(0.32)
Weib/Weib	(0.315)	0.367(0.34)	0.370(0.36)	0.335(0.46)	0.1435(0.35)	0.300(0.33)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.324(0.32)	0.360(0.32)	0.292(0.42)	0.075(0.34)	0.231(0.31)
Exp/Weib	(0.526)	0.172(0.26)	0.180(0.27)	0.171(0.30)	0.040(0.30)	0.135(0.26)
Weib/Exp	(0.315)	0.427(0.32)	0.430(0.33)	0.346(0.43)	0.152(0.34)	0.339(0.31)
Weib/Weib	(0.315)	0.391(0.33)	0.420(0.34)	0.349(0.45)	0.143(0.36)	0.324(0.32)

Tabla 2.5: TE y LPI de NN 95 % con 45 % de censura para $n = 25$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.056(0.26)	0.040(0.27)	0.017(0.42)	0.082(0.26)	0.056(0.26)
Exp/Weib	(0.526)	0.043(0.22)	0.060(0.22)	0.015(0.28)	0.066(0.22)	0.075(0.22)
Weib/Exp	(0.315)	0.084(0.26)	0.130(0.27)	0.033(0.41)	0.033(0.26)	0.050(0.26)
Weib/Weib	(0.315)	0.084(0.26)	0.130(0.27)	0.033(0.41)	0.033(0.26)	0.050(0.26)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.056(0.26)	0.060(0.26)	0.017(0.41)	0.048(0.26)	0.056(0.26)
Exp/Weib	(0.526)	0.043(0.22)	0.070(0.21)	0.015(0.28)	0.066(0.23)	0.043(0.21)
Weib/Exp	(0.315)	0.076(0.26)	0.120(0.26)	0.033(0.41)	0.033(0.26)	0.042(0.26)
Weib/Weib	(0.315)	0.076(0.26)	0.120(0.26)	0.033(0.41)	0.033(0.26)	0.042(0.26)

Tabla 2.6: TE y LPI de NN 95 % con 0 % de censura para $n = 50$

2.2.2. Resultados para $n = 50$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.091(0.25)	0.100(0.26)	0.057(0.37)	0.045(0.25)	0.082(0.25)
Exp/Weib	(0.526)	0.073(0.21)	0.100(0.21)	0.065(0.26)	0.038(0.22)	0.074(0.21)
Weib/Exp	(0.315)	0.184(0.25)	0.250(0.26)	0.124(0.37)	0.079(0.25)	0.153(0.25)
Weib/Weib	(0.315)	0.163(0.26)	0.210(0.26)	0.104(0.38)	0.085(0.26)	0.146(0.25)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.091(0.25)	0.080(0.25)	0.069(0.37)	0.045(0.26)	0.090(0.25)
Exp/Weib	(0.526)	0.073(0.21)	0.130(0.21)	0.065(0.25)	0.035(0.22)	0.074(0.20)
Weib/Exp	(0.315)	0.211(0.25)	0.280(0.25)	0.139(0.37)	0.079(0.26)	0.154(0.25)
Weib/Weib	(0.315)	0.180(0.25)	0.220(0.25)	0.115(0.37)	0.085(0.26)	0.152(0.25)

Tabla 2.7: TE y LPI de NN 95 % con 15 % de censura para $n = 50$

En el segundo escenario se presenta el mismo patrón de comportamiento que cuando

tenemos muestras de tamaño $n = 25$, es decir, que cuando no hay observaciones censuradas los I.C mediante la transformación $\log(S(t))$ resultan ser mejores.

Por otra parte a medida que se aumenta el porcentaje de censura, los I.C mediante la transformación $\log(-\log(S(t)))$ son más efectivos, en particular utilizando el estimador de K.M y mediante la distribución de tiempos/censura Weib/Exp.

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.172(0.25)	0.204(0.25)	0.110(0.35)	0.084(0.25)	0.145(0.24)
Exp/Weib	(0.526)	0.113(0.20)	0.168(0.21)	0.111(0.25)	0.033(0.21)	0.097(0.20)
Weib/Exp	(0.315)	0.293(0.25)	0.260(0.24)	0.205(0.35)	0.161(0.25)	0.260(0.24)
Weib/Weib	(0.315)	0.270(0.25)	0.265(0.25)	0.190(0.36)	0.158(0.25)	0.234(0.25)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.174(0.25)	0.234(0.24)	0.121(0.34)	0.086(0.25)	0.157(0.24)
Exp/Weib	(0.526)	0.113(0.20)	0.174(0.20)	0.111(0.24)	0.030(0.21)	0.104(0.20)
Weib/Exp	(0.315)	0.321(0.25)	0.250(0.30)	0.228(0.35)	0.162(0.25)	0.265(0.24)
Weib/Weib	(0.315)	0.287(0.25)	0.253(0.31)	0.205(0.35)	0.162(0.26)	0.251(0.25)

Tabla 2.8: TE y LPI de NN 95 % con 25 % de censura para $n = 50$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.389(0.24)	0.430(0.24)	0.276(0.31)	0.230(0.24)	0.321(0.23)
Exp/Weib	(0.526)	0.222(0.19)	0.246(0.20)	0.216(0.23)	0.053(0.21)	0.150(0.19)
Weib/Exp	(0.315)	0.537(0.24)	0.341(0.22)	0.417(0.32)	0.371(0.24)	0.498(0.23)
Weib/Weib	(0.315)	0.671(0.24)	0.500(0.21)	0.549(0.31)	0.480(0.24)	0.606(0.23)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.399(0.23)	0.450(0.23)	0.291(0.31)	0.236(0.24)	0.352(0.23)
Exp/Weib	(0.526)	0.224(0.19)	0.256(0.19)	0.219(0.22)	0.053(0.21)	0.167(0.19)
Weib/Exp	(0.315)	0.561(0.24)	0.443(0.20)	0.444(0.31)	0.377(0.24)	0.513(0.23)
Weib/Weib	(0.315)	0.690(0.23)	0.556(0.21)	0.565(0.30)	0.491(0.24)	0.632(0.23)

Tabla 2.9: TE y LPI de NN 95 % con 35 % de censura para $n = 50$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.576(0.23)	0.590(0.24)	0.457(0.29)	0.388(0.23)	0.500(0.22)
Exp/Weib	(0.526)	0.333(0.18)	0.310(0.19)	0.323(0.21)	0.088(0.20)	0.225(0.18)
Weib/Exp	(0.315)	0.731(0.23)	0.800(0.23)	0.616(0.29)	0.565(0.23)	0.685(0.22)
Weib/Weib	(0.315)	0.671(0.24)	0.740(0.24)	0.549(0.31)	0.480(0.24)	0.606(0.23)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.593(0.23)	0.600(0.23)	0.474(0.29)	0.395(0.23)	0.527(0.22)
Exp/Weib	(0.526)	0.337(0.18)	0.320(0.19)	0.330(0.21)	0.088(0.20)	0.245(0.18)
Weib/Exp	(0.315)	0.744(0.23)	0.810(0.22)	0.645(0.29)	0.571(0.23)	0.705(0.22)
Weib/Weib	(0.315)	0.690(0.23)	0.740(0.24)	0.565(0.30)	0.491(0.24)	0.632(0.23)

Tabla 2.10: TE y LPI de NN 95 % con 45 % de censura para $n = 50$

2.2.3. Resultados para $n = 75$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.050(0.21)	0.064(0.21)	0.008(0.33)	0.055(0.21)	0.068(0.21)
Exp/Weib	(0.526)	0.053(0.18)	0.074(0.18)	0.012(0.22)	0.049(0.18)	0.055(0.18)
Weib/Exp	(0.315)	0.077(0.21)	0.080(0.22)	0.040(0.33)	0.059(0.21)	0.077(0.21)
Weib/Weib	(0.315)	0.077(0.21)	0.075(0.23)	0.040(0.33)	0.059(0.21)	0.077(0.21)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.050(0.21)	0.074(0.21)	0.008(0.33)	0.055(0.21)	0.050(0.21)
Exp/Weib	(0.526)	0.053(0.18)	0.084(0.17)	0.029(0.22)	0.049(0.18)	0.055(0.18)
Weib/Exp	(0.315)	0.077(0.21)	0.066(0.20)	0.040(0.33)	0.050(0.21)	0.077(0.21)
Weib/Weib	(0.315)	0.077(0.21)	0.055(0.20)	0.040(0.33)	0.050(0.21)	0.077(0.21)

Tabla 2.11: TE y LPI de NN 95 % con 0 % de censura para $n = 75$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.115(0.21)	0.101(0.23)	0.064(0.30)	0.065(0.21)	0.087(0.21)
Exp/Weib	(0.526)	0.093(0.17)	0.095(0.20)	0.060(0.21)	0.037(0.18)	0.066(0.17)
Weib/Exp	(0.315)	0.261(0.21)	0.112(0.24)	0.144(0.30)	0.174(0.21)	0.221(0.21)
Weib/Weib	(0.315)	0.233(0.21)	0.122(0.24)	0.128(0.31)	0.150(0.21)	0.202(0.21)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.119(0.21)	0.124(0.20)	0.068(0.30)	0.067(0.21)	0.102(0.20)
Exp/Weib	(0.526)	0.100(0.17)	0.110(0.21)	0.065(0.21)	0.036(0.18)	0.067(0.17)
Weib/Exp	(0.315)	0.263(0.21)	0.120(0.21)	0.161(0.30)	0.181(0.21)	0.245(0.20)
Weib/Weib	(0.315)	0.249(0.21)	0.120(0.23)	0.139(0.30)	0.157(0.21)	0.218(0.21)

Tabla 2.12: TE y LPI de NN 95 % con 15 % de censura para $n = 75$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.279(0.20)	0.162(0.23)	0.170(0.28)	0.162(0.20)	0.235(0.20)
Exp/Weib	(0.526)	0.153(0.16)	0.156(0.24)	0.136(0.20)	0.065(0.17)	0.129(0.16)
Weib/Exp	(0.315)	0.493(0.20)	0.255(0.23)	0.337(0.28)	0.371(0.20)	0.450(0.20)
Weib/Weib	(0.315)	0.440(0.21)	0.255(0.23)	0.303(0.28)	0.325(0.21)	0.395(0.20)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.293(0.20)	0.197(0.24)	0.189(0.27)	0.170(0.20)	0.249(0.20)
Exp/Weib	(0.526)	0.168(0.16)	0.189(0.24)	0.137(0.19)	0.065(0.17)	0.135(0.16)
Weib/Exp	(0.315)	0.513(0.20)	0.354(0.25)	0.356(0.27)	0.382(0.20)	0.469(0.20)
Weib/Weib	(0.315)	0.461(0.20)	0.361(0.24)	0.319(0.28)	0.331(0.21)	0.413(0.20)

Tabla 2.13: TE y LPI de NN 95 % con 25 % de censura para $n = 75$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.506(0.20)	0.391(0.22)	0.380(0.26)	0.369(0.20)	0.461(0.20)
Exp/Weib	(0.526)	0.268(0.16)	0.314(0.15)	0.254(0.18)	0.117(0.16)	0.227(0.16)
Weib/Exp	(0.315)	0.725(0.20)	0.322(0.16)	0.580(0.26)	0.605(0.20)	0.686(0.20)
Weib/Weib	(0.315)	0.665(0.20)	0.322(0.15)	0.530(0.27)	0.544(0.20)	0.640(0.20)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.525(0.20)	0.395(0.19)	0.396(0.25)	0.376(0.20)	0.470(0.19)
Exp/Weib	(0.526)	0.273(0.16)	0.321(0.19)	0.258(0.18)	0.119(0.17)	0.234(0.16)
Weib/Exp	(0.315)	0.740(0.20)	0.315(0.21)	0.593(0.25)	0.621(0.20)	0.702(0.21)
Weib/Weib	(0.315)	0.680(0.20)	0.320(0.22)	0.542(0.26)	0.555(0.20)	0.642(0.20)

Tabla 2.14: TE y LPI de NN 95 % con 35 % de censura para $n = 75$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.797(0.18)	0.562(0.21)	0.681(0.23)	0.670(0.19)	0.755(0.18)
Exp/Weib	(0.526)	0.452(0.15)	0.553(0.15)	0.427(0.17)	0.234(0.16)	0.376(0.15)
Weib/Exp	(0.315)	0.921(0.18)	0.601(0.19)	0.843(0.23)	0.856(0.19)	0.902(0.18)
Weib/Weib	(0.315)	0.870(0.19)	0.589(0.20)	0.779(0.24)	0.791(0.19)	0.851(0.19)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.806(0.18)	0.661(0.22)	0.693(0.23)	0.678(0.19)	0.768(0.18)
Exp/Weib	(0.526)	0.459(0.15)	0.532(0.16)	0.438(0.17)	0.235(0.16)	0.390(0.15)
Weib/Exp	(0.315)	0.927(0.18)	0.691(0.21)	0.852(0.23)	0.861(0.19)	0.908(0.18)
Weib/Weib	(0.315)	0.878(0.19)	0.610(0.21)	0.791(0.24)	0.799(0.19)	0.860(0.19)

Tabla 2.15: TE y LPI de NN 95 % con 45 % de censura para $n = 75$

De igual manera los resultados para los escenarios anteriores respecto a que cuando no se presentan observaciones censuradas en los datos, los intervalos de confianza mediante la transformación $\log(S(t))$ son más atractivos, como también el hecho de que el esti-

mador de Kaplan-Meier mediante la transformación $\log(-\log(S(t)))$ presenta mejores resultados a medida que se aumenta el porcentaje de censura.

Es de resaltar que los intervalos de confianza para la función de supervivencia mediante el remuestreo *Bootstrap* a los estimadores de Kaplan-Meier y Nelson-Aalen arrojan resultados inapropiados, esto se debe a que el tipo que se utilizó fue normal, el cual es sensible a distribuciones asimétricas.

2.2.4. Resultados para $n = 100$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.056(0.18)	0.040(0.18)	0.006(0.28)	0.060(0.18)	0.056(0.18)
Exp/Weib	(0.526)	0.053(0.15)	0.070(0.15)	0.021(0.19)	0.064(0.16)	0.053(0.15)
Weib/Exp	(0.315)	0.090(0.19)	0.190(0.18)	0.037(0.28)	0.062(0.18)	0.062(0.18)
Weib/Weib	(0.315)	0.090(0.19)	0.190(0.18)	0.037(0.28)	0.062(0.18)	0.062(0.18)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.056(0.18)	0.040(0.18)	0.013(0.28)	0.044(0.19)	0.056(0.18)
Exp/Weib	(0.526)	0.053(0.15)	0.070(0.15)	0.021(0.19)	0.064(0.16)	0.053(0.15)
Weib/Exp	(0.315)	0.085(0.18)	0.180(0.18)	0.037(0.28)	0.062(0.19)	0.090(0.18)
Weib/Weib	(0.315)	0.085(0.18)	0.180(0.18)	0.037(0.28)	0.062(0.19)	0.090(0.18)

Tabla 2.16: TE y LPI de NN 95 % con 0 % de censura para $n = 100$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.093(0.15)	0.240(0.18)	0.086(0.18)	0.047(0.15)	0.080(0.15)
Exp/Weib	(0.526)	0.094(0.15)	0.100(0.15)	0.076(0.18)	0.041(0.15)	0.070(0.15)
Weib/Exp	(0.315)	0.330(0.18)	0.400(0.18)	0.195(0.26)	0.239(0.18)	0.299(0.18)
Weib/Weib	(0.315)	0.297(0.18)	0.380(0.18)	0.170(0.26)	0.208(0.18)	0.271(0.18)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.093(0.15)	0.250(0.18)	0.089(0.17)	0.048(0.15)	0.086(0.14)
Exp/Weib	(0.526)	0.095(0.15)	0.100(0.15)	0.081(0.18)	0.039(0.15)	0.075(0.15)
Weib/Exp	(0.315)	0.341(0.18)	0.420(0.18)	0.203(0.26)	0.249(0.18)	0.311(0.18)
Weib/Weib	(0.315)	0.311(0.18)	0.420(0.18)	0.178(0.26)	0.216(0.18)	0.283(0.18)

Tabla 2.17: TE y LPI de NN 95 % con 15 % de censura para $n = 100$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.355(0.17)	0.231(0.19)	0.229(0.24)	0.255(0.18)	0.314(0.17)
Exp/Weib	(0.526)	0.193(0.14)	0.231(0.15)	0.144(0.17)	0.085(0.15)	0.139(0.14)
Weib/Exp	(0.315)	0.598(0.17)	0.504(0.17)	0.436(0.24)	0.502(0.18)	0.572(0.17)
Weib/Weib	(0.315)	0.542(0.18)	0.561(0.19)	0.383(0.25)	0.450(0.18)	0.510(0.18)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.367(0.17)	0.334(0.20)	0.237(0.24)	0.266(0.18)	0.323(0.17)
Exp/Weib	(0.526)	0.200(0.14)	0.212(0.21)	0.149(0.17)	0.086(0.15)	0.141(0.14)
Weib/Exp	(0.315)	0.617(0.17)	0.617(0.22)	0.449(0.24)	0.515(0.18)	0.585(0.17)
Weib/Weib	(0.315)	0.556(0.18)	0.618(0.22)	0.395(0.24)	0.460(0.18)	0.525(0.18)

Tabla 2.18: TE y LPI de NN 95 % con 25 % de censura para $n = 100$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.666(0.17)	0.675(0.18)	0.509(0.22)	0.548(0.17)	0.626(0.17)
Exp/Weib	(0.526)	0.336(0.14)	0.321(0.15)	0.294(0.16)	0.183(0.14)	0.280(0.14)
Weib/Exp	(0.315)	0.857(0.17)	0.657(0.18)	0.730(0.22)	0.788(0.17)	0.838(0.17)
Weib/Weib	(0.315)	0.789(0.17)	0.699(0.18)	0.657(0.23)	0.714(0.17)	0.764(0.17)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.679(0.17)	0.581(0.19)	0.526(0.22)	0.557(0.17)	0.639(0.17)
Exp/Weib	(0.526)	0.348(0.14)	0.299(0.19)	0.297(0.16)	0.185(0.14)	0.286(0.13)
Weib/Exp	(0.315)	0.867(0.17)	0.712(0.17)	0.740(0.22)	0.798(0.17)	0.849(0.17)
Weib/Weib	(0.315)	0.801(0.17)	0.701(0.17)	0.671(0.23)	0.721(0.17)	0.775(0.17)

Tabla 2.19: TE y LPI de NN 95 % con 35 % de censura para $n = 100$

Estimador $\hat{S}_{KM}(0.2)$						
Time/Cens	LI.PAR	T.KM	Boot.KM	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.890(0.16)	0.890(0.16)	0.804(0.20)	0.826(0.16)	0.874(0.16)
Exp/Weib	(0.526)	0.534(0.13)	0.540(0.13)	0.497(0.15)	0.351(0.13)	0.475(0.13)
Weib/Exp	(0.315)	0.977(0.16)	0.940(0.16)	0.930(0.20)	0.957(0.16)	0.970(0.16)
Weib/Weib	(0.315)	0.940(0.17)	0.870(0.17)	0.880(0.21)	0.909(0.17)	0.932(0.16)
Estimador $\hat{S}_{NA}(0.2)$						
Time/Cens	LI.PAR	T.NA	Boot.NA	log	log(-log)	Arco seno
Exp/Exp	(0.526)	0.895(0.16)	0.910(0.16)	0.813(0.20)	0.833(0.16)	0.881(0.16)
Exp/Weib	(0.526)	0.543(0.13)	0.560(0.13)	0.506(0.15)	0.354(0.13)	0.483(0.13)
Weib/Exp	(0.315)	0.980(0.16)	0.930(0.16)	0.935(0.20)	0.958(0.16)	0.975(0.16)
Weib/Weib	(0.315)	0.951(0.16)	0.870(0.16)	0.888(0.21)	0.911(0.17)	0.935(0.16)

Tabla 2.20: TE y LPI de NN 95 % con 45 % de censura para $n = 100$

Para el último escenario la comparación de estos intervalos de confianza para la función de supervivencia en muestras de tamaño $n = 100$, los resultados respaldan lo mencionado con los demás tamaños de muestra, sin embargo es de resaltar que a medida que se

aumenta el porcentaje de censura al caso más extremo, las tasas de error son bastantes altas en particular al aumentar de 35 % a 45 % observaciones censuradas en la muestra simulada.

Teniendo en cuenta los resultados anteriores en algunos casos se presentan confusiones para la escogencia del IC que presenta mejores resultados por lo que se implementó la metodología de Correa y Sierra (2003) para comparar dichos intervalos a través del índice propuesto por dichos autores, modificando el nivel de confianza nominal ($1-\alpha=0.9, 0.95$ y 0.99), lo anterior se muestra tomando la distribución de fallas/censura exp/exp, esto con el fin de presentar el funcionamiento del índice.

Es de resaltar que mientras mayor sea el índice mejor será el intervalo de confianza, sin embargo éste índice fue propuesto para comparar intervalos de confianza para diferencia de proporciones, lo que cambiaría el rango de valores resultantes, pero la analogía de la interpretación se mantiene.

NN	Trad.K-M	Boot.K-M	log	log(-log)	Arco seno
n=25					
90 %	0.8466671	0.84123521	0.8701985	0.8894355	0.8438988
95 %	0.7821166	0.77833918	0.7926120	0.8123433	0.7951308
99 %	0.7258416	0.72728372	0.7269520	0.7365756	0.7302833
n=50					
90 %	0.8267704	0.85635273	0.8939118	0.8996994	0.8424531
95 %	0.8223925	0.82356248	0.8536054	0.8644621	0.8305350
99 %	0.7928487	0.79263571	0.8080810	0.8179597	0.8076683
n=75					
90 %	0.8205188	0.84526372	0.8973483	0.8762578	0.8555834
95 %	0.8248394	0.81425311	0.8728385	0.8719065	0.8509360
99 %	0.8244178	0.82416387	0.8390008	0.8497242	0.8390008
n=100					
90 %	0.8933075	0.82414272	0.8947079	0.8529202	0.8506308
95 %	0.8375906	0.82854771	0.8445166	0.8800704	0.8495958
99 %	0.8274756	0.82251437	0.8342818	0.8643899	0.8428562

Tabla 2.21: Índice de IC para $\hat{S}_{KM}(t)$ con 15 % de censura

Entonces tenemos que los resultados para el estimador de Kaplan-Meier muestran que en realidad los intervalos de confianza mediante la transformación $\log(-\log(S(t)))$ son más atractivos cuando tenemos observaciones censuradas en particular este caso con 15 % de censura. También vemos que el índice está afectado por el nivel de confianza ya que a medida que aumenta NN el índice es menor.

NN	Trad.N-A	Boot.N-A	log	log(-log)	Arco seno
n=25					
90 %	0.7982203	0.81425314	0.8701985	0.8751231	0.8508197
95 %	0.7842156	0.77270297	0.7921921	0.8140226	0.7884138
99 %	0.7258416	0.75271438	0.7269521	0.7391666	0.7295439
n=50					
90 %	0.8282407	0.85647321	0.8939118	0.8946399	0.8346117
95 %	0.8228449	0.84649534	0.8422964	0.8644621	0.8237496
99 %	0.7903788	0.79627382	0.8080873	0.8183713	0.8076683
n=75					
90 %	0.8154972	0.85647262	0.8908203	0.8977643	0.8240339
95 %	0.8211113	0.81657243	0.8691104	0.8695765	0.8374217
99 %	0.8231301	0.81627352	0.8390008	0.8497242	0.8342825
n=100					
90 %	0.8792314	0.79534263	0.8845098	0.8847683	0.7978937
95 %	0.8375906	0.73974607	0.8512845	0.8791469	0.8445166
99 %	0.8239657	0.70627163	0.8438423	0.8635101	0.8415379

Tabla 2.22: Índice de IC para $\hat{S}_{NA}(t)$ con 15 % de censura

De igual forma el estimador de Nelson-Aalen mediante la estimación de la varianza de Aalen y Johansen los intervalos de confianza mediante la transformación $\log(-\log(S(t)))$ son más eficientes que los demás.

Capítulo 3

CONCLUSIONES

- Cuando no se presentan observaciones censuradas en los datos los IC mediante la transformación \log poseen menor tasas de error, independientemente de la distribución de tiempo/censura y tamaños de muestra, como también el estimador utilizado, esto se debe a que resultan ser más amplios.
- A medida que se aumenta el porcentaje de censura, los IC mediante la transformación $\log(-\log)$ resultan ser más efectivos, independientemente de los modelos generadores, tamaños de muestra y estimador de supervivencia utilizado.
- El estimador de KM presenta mejor comportamiento que el estimador NA, a medida que se aumenta el tamaño de muestra y porcentajes de censura, independientemente de la distribución de la falla/censura.
- Cuando el tiempo de falla/censura se distribuyen exp/exp los resultados de la estimación de $S(0.2)$ resultan ser mejores que las demás combinaciones de distribución de falla/censura, en particular utilizando el estimador de KM.
- El estimador de NA es más efectivo que el estimador KM, solo en el caso de muestras de tamaño $n = 25$ y tiempos distribuidos Weibull, con datos sin censurar.
- Los intervalos de confianza no paramétricos frente a los paramétricos resultan ser mejores mientras el tamaño de muestra se aumente, en particular cuando se utiliza la transformación $\log(-\log(S(t)))$.
- La propuesta de Correa y Sierra (2003); del índice para comparar intervalos de confianza para proporciones, puede ser utilizado para comparar intervalos de confianza para a función de supervivencia en un tiempo de interés con censura a derecha.

Capítulo 4

REFERENCIAS

1. Aalen, O. & Johansen, S. (1978). An Empirical Transition Matrix for Nonhomogeneous Markov Chains Based on Censored Observations. *Scandinavian Journal of Statistics*. Vol. 5, No. 3, pp. 141-150.
2. Akritas, M. (1986). Bootstrapping the Kaplan-Meier Estimator. *American Statistical Association* Vol. 81, No. 396, pp. 1032- 1038.
3. Anderson, J. Bernstein, L. & Pike, M. (1982). Confidence Intervals for Probabilities of Survival and Quantiles in Life-Table Analysis. *Biometrics* Vol. 38, No. 2, pp. 407- 416.
4. Barber, S. & Jennison, C. (1999). Symmetric Tests Confidence Intervals for Survival Probabilities and Quantiles of Censored Data. *Biometrics* Vol. 55, No. 2, pp. 430- 436.
5. Bie, O. Borgan, Ø. & Liestøl, K. (1987). Confidence Intervals and confidence Bands for the Cumulative Hazard Rate Function and Their Small Sample Properties. *Scandinavian Journal of Statistics* Vol. 14, No. 3, pp. 221- 233.
6. Borgan, Ø & Liestøl, K. (1990). A Note on Confidence Intervals and Bands for the Survival Function Based on Transformations. *Scandinavian Journal of Statistics* Vol. 17, No. 1, pp. 35-41.
7. Brookmeyer, R & Crowley, J.(1982). A Confidence Interval for the Median Survival Time. *Biometrics*. Vol. 38, No. 1, pp. 29-41.
8. Correa, J & Sierra, E. (2003). Intervalos de confianza para la comparación de dos proporciones. *Revista Colombiana de Estadística* Vol. 26, No. 1, pp. 61-75.

9. DiCiccio, T. & Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science* Vol. 11, No. 3, pp. 189-212.
10. Emerson, J. (1982). Nonparametric Confidence Intervals for the Median in the Presence of Right Censoring. *Biometrics* Vol. 38, No. 1, pp. 17-27.
11. Efron, B & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* Vol. 1, No. 1, pp. 54-75.
12. Greenwood, M. (1926). The natural duration of cancer. *Reports on Public Health and Medical Subjects*. No. 33, pp. 1-26.
13. Jeng, Shuen-Lin & Meeker, W. (2000). Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data. *Technometrics*. Vol. 42, No. 2, pp. 135-148.
14. Jennison, C & Turnbull, B. (1985). Repeated Confidence Intervals for the Median Survival Time. *Biometrika* Vol. 70, No. 3, pp. 619-625.
15. Kalbfleisch, J & Prentice, R. (1980). The statistical analysis of failure time data. *Wiley* New York.
16. Kaplan, E & Meier, P. (1958). Estimation from Incomplete Observations. *American Statistical Association* Vol. 53, No. 282, pp. 457- 481.
17. Lawless, J. (2003). *Statistical Models and Methods For Lifetime Data*. John Wiley and Sons.
18. Molinero, L (2001). Tiempo hasta que ocurre un suceso. *Asociación de la sociedad Española de Hipertensión*. Junio 2001.
19. Nair, N. (1993). Bounds for Reliability Estimation under Dependent Censoring. *International Statistical Review*. Vol. 61, No. 1, pp. 169-182.
20. Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*. Vol. 61, No. 1, pp. 27-52.
21. Reid, N. (1981). Estimating the Median Survival Time. *Biometrika*. Vol. 68, No. 3, pp. 601-608.
22. Selvin, S. (2008). *Survival Analysis for Epidemiologic and Medical Research*. *Cambridge University Press* New York.
23. Slud, E. Byar, D. & Green, S. (1984). A Comparison of Reflected Versus Test-Based Confidence Intervals for the Median Survival Time, Based on Censored Data. *Biometrics*. Vol. 40, No. 3, pp. 587-600.

24. Tomas, D. & Grunkemeier, G. (1975). Confidence Interval Estimation of Survival Probabilities for Censored Data. *Journal of American Statistical Association*. Vol. 70, No. 352, pp. 865-871.

Capítulo 5

ANEXOS

5.1. Graficos

Teniendo en cuenta los resultados presentados en las tablas del capítulo anterior se realizan los respectivos gráficos, con el fin de ilustrar los comportamientos y tendencias de las tasas de error de los intervalos de confianza para la función de supervivencia en un tiempo de interés con censura a derecha, utilizando estimadores no paramétricos de Kaplan-Meier y Nelson-Aalen, de una manera fácil, para que el lector pueda sacar sus propias conclusiones.

Sin pérdida de generalidad las abreviaturas corresponden a: T.KM: I.C mediante el estimador de Kaplan-Meier, Boot.KM y Boot.NA: I.C mediante el remuestreo *Bootstrap* a los estimadores de Kaplan-Meier y Nelson-Aalen respectivamente, además \log , $\log(-\log)$ y *Arco seno* corresponden a las transformaciones de las funciones de supervivencia definidas en (1.42), (1.45) y (1.46) respectivamente.

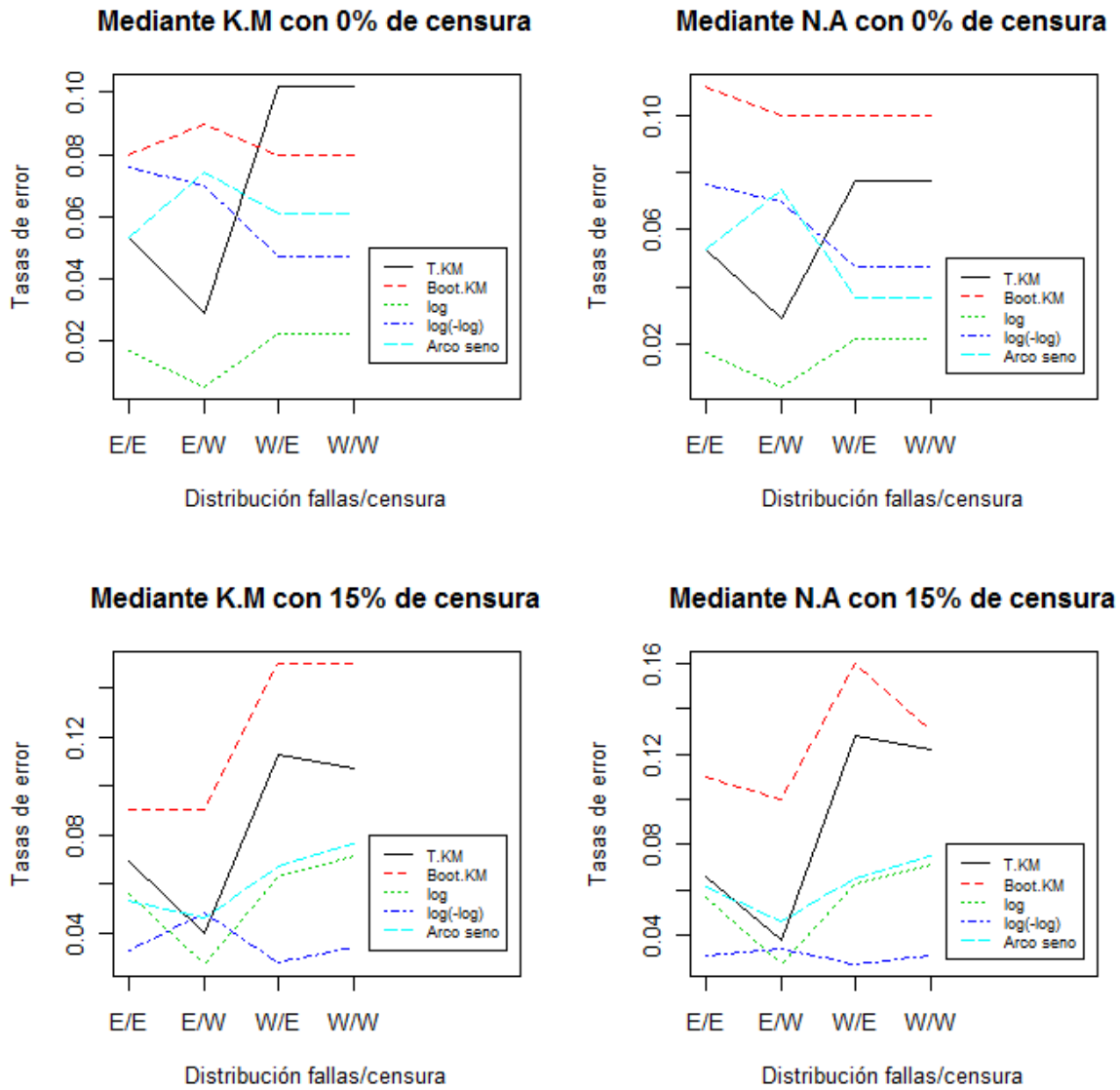


Figura 5.1: Tasas de Error de I.C para $S(t)$ con $n = 25$

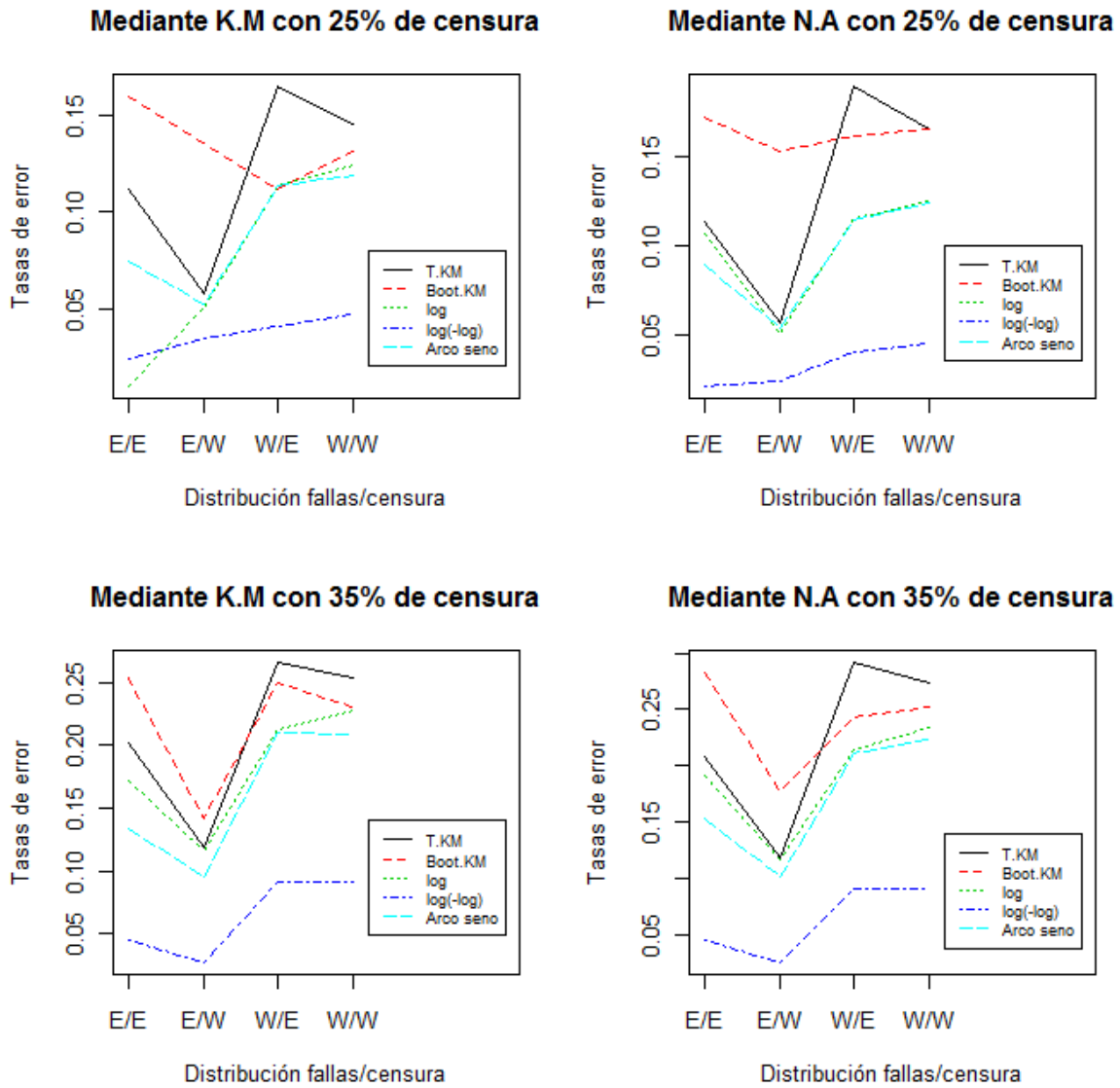


Figura 5.2: Tasas de Error de I.C para $S(t)$ con $n = 25$

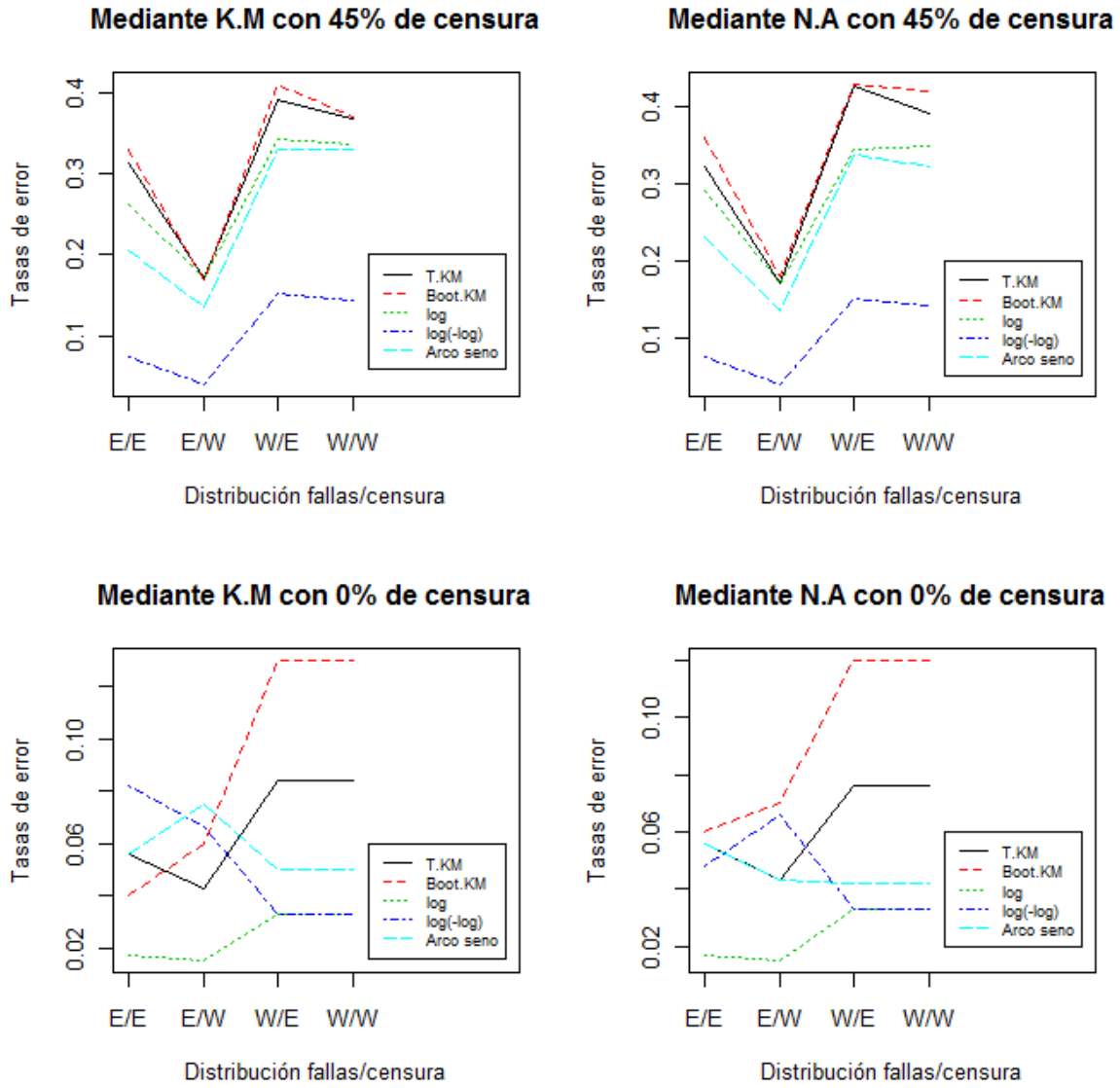


Figura 5.3: Tasas de Error de I.C para $S(t)$ con $n = 25$ y $n = 50$

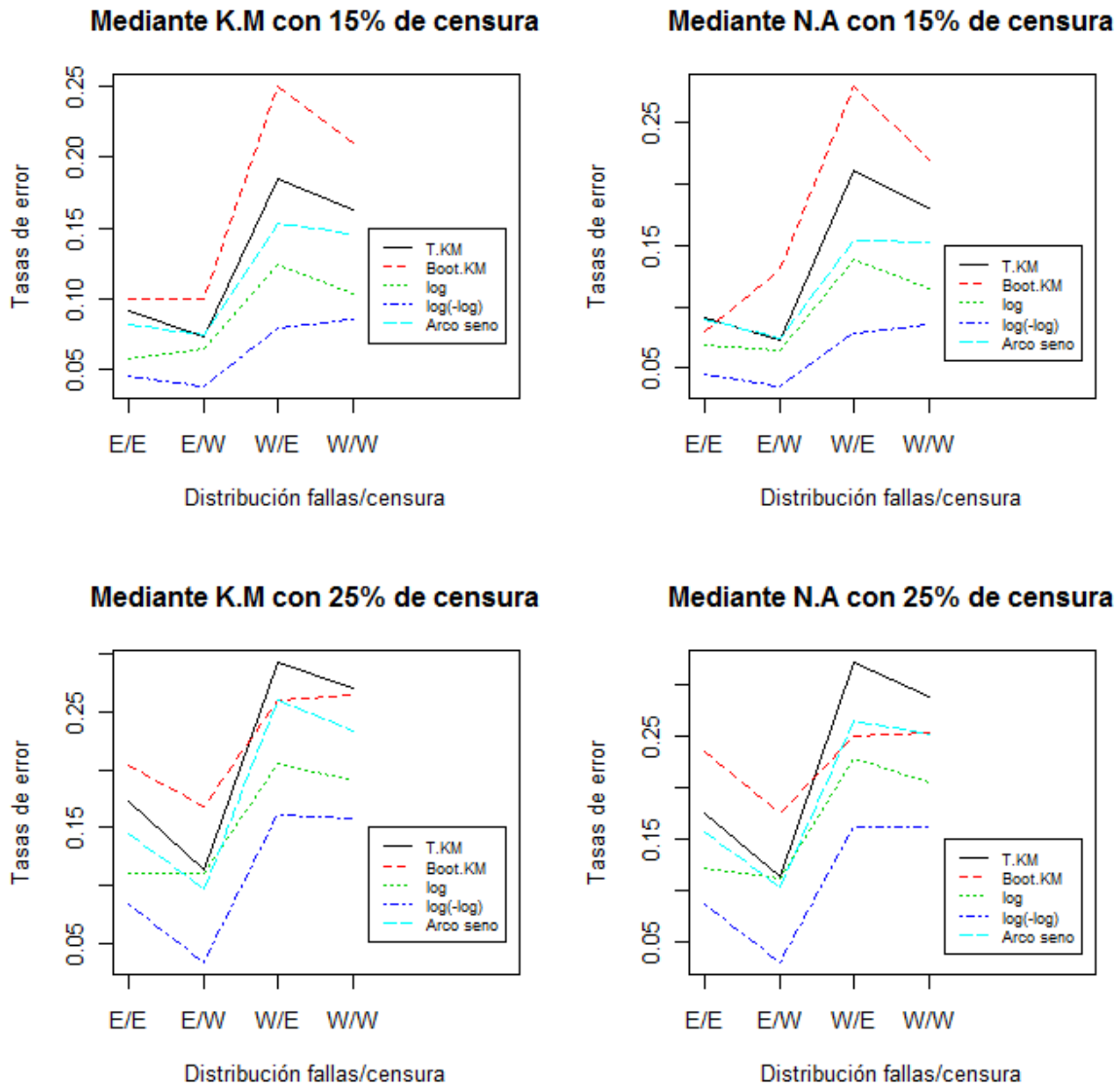


Figura 5.4: Tasas de Error de I.C para $S(t)$ con $n = 50$

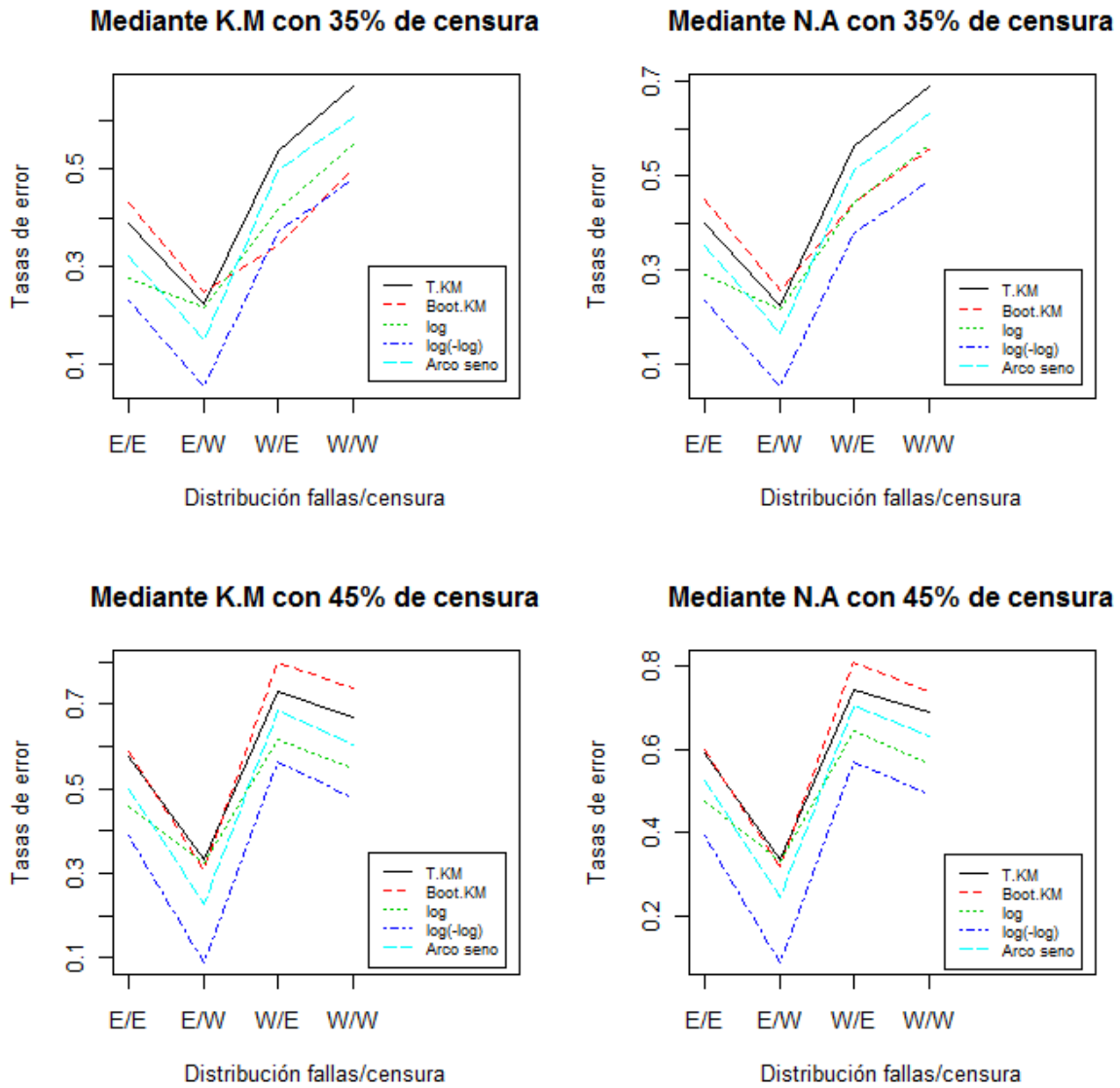


Figura 5.5: Tasas de Error de I.C para $S(t)$ con $n = 50$

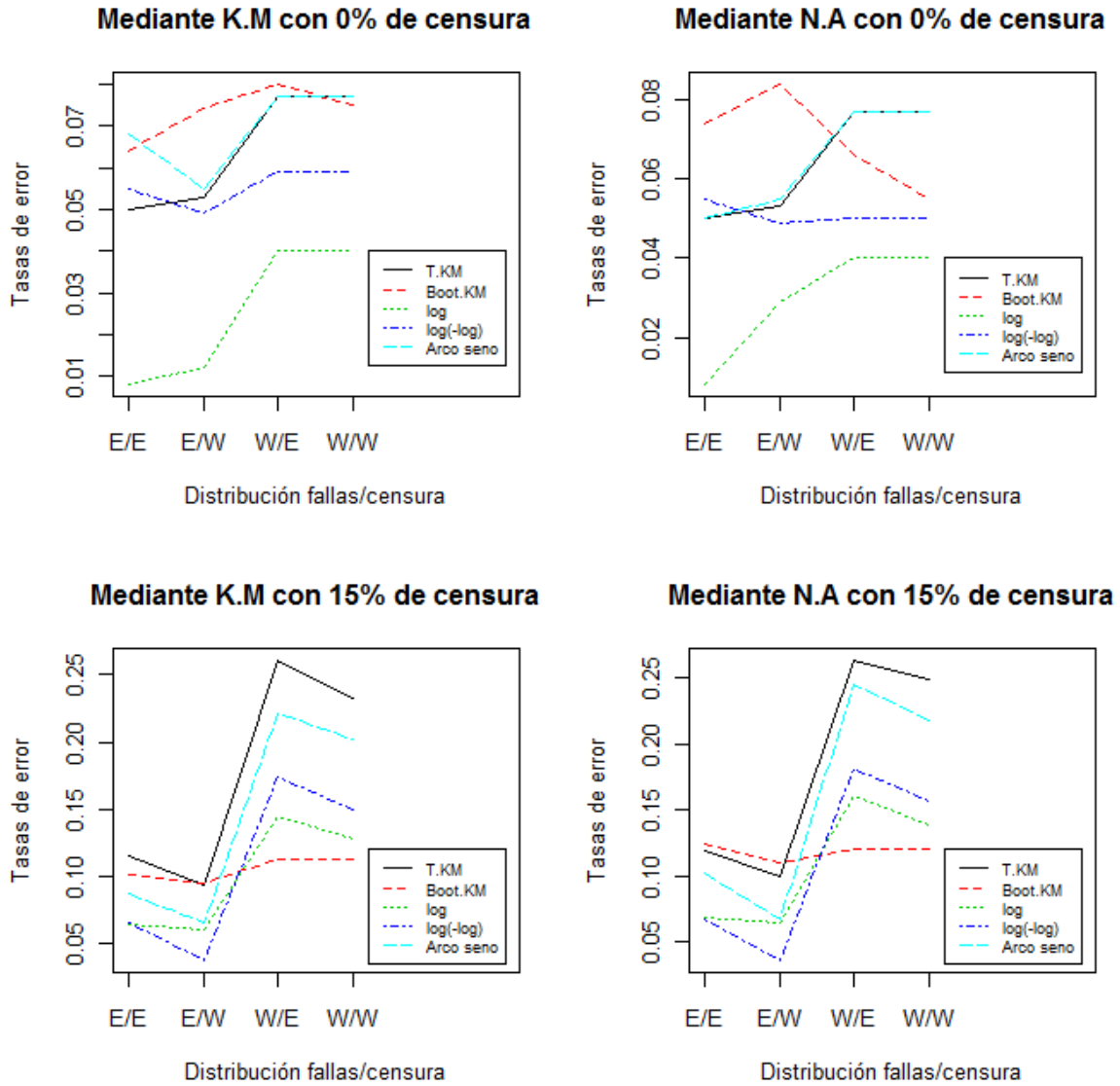


Figura 5.6: Tasas de Error de I.C para $S(t)$ con $n = 75$

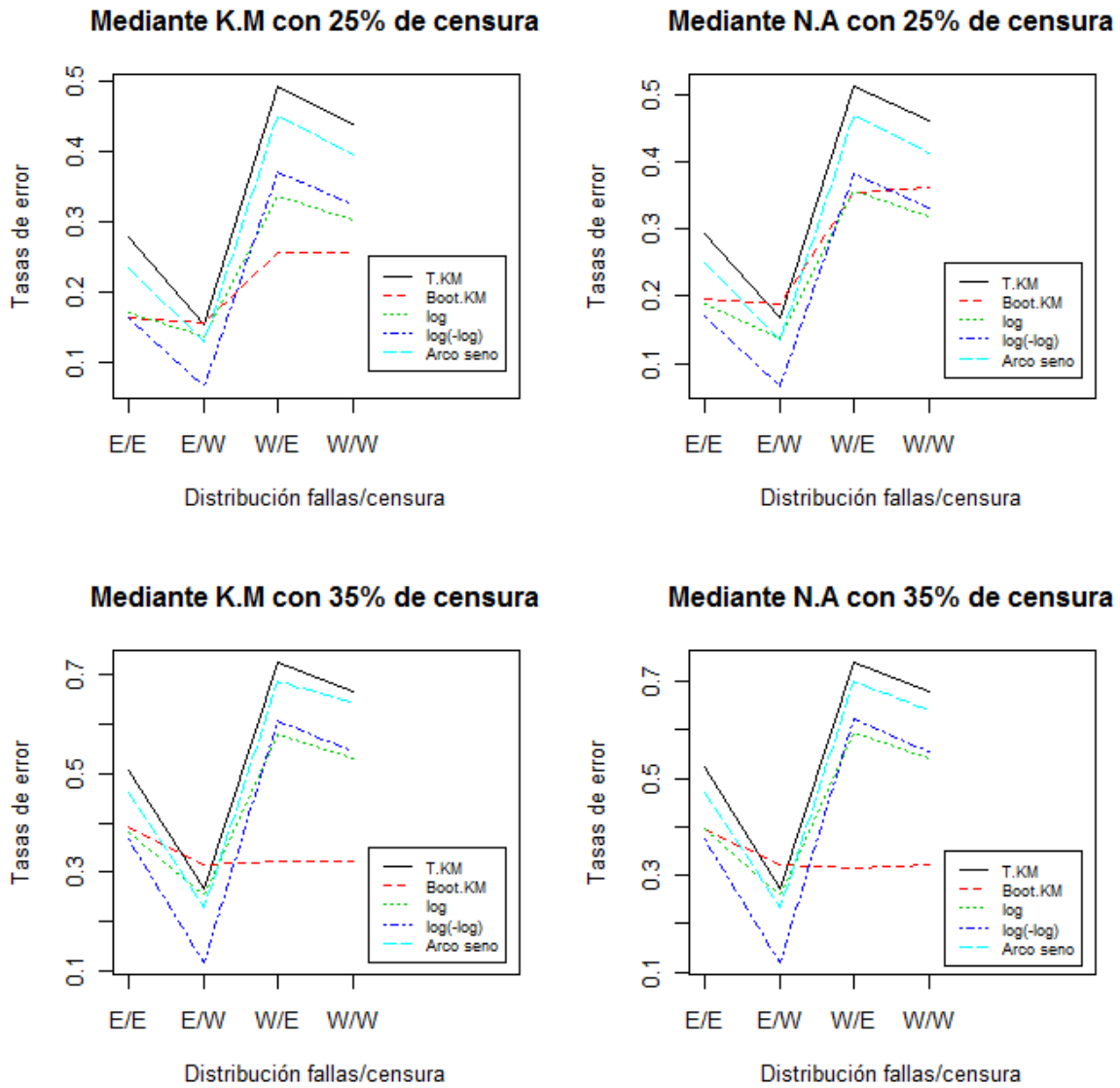


Figura 5.7: Tasas de Error de I.C para $S(t)$ con $n = 75$

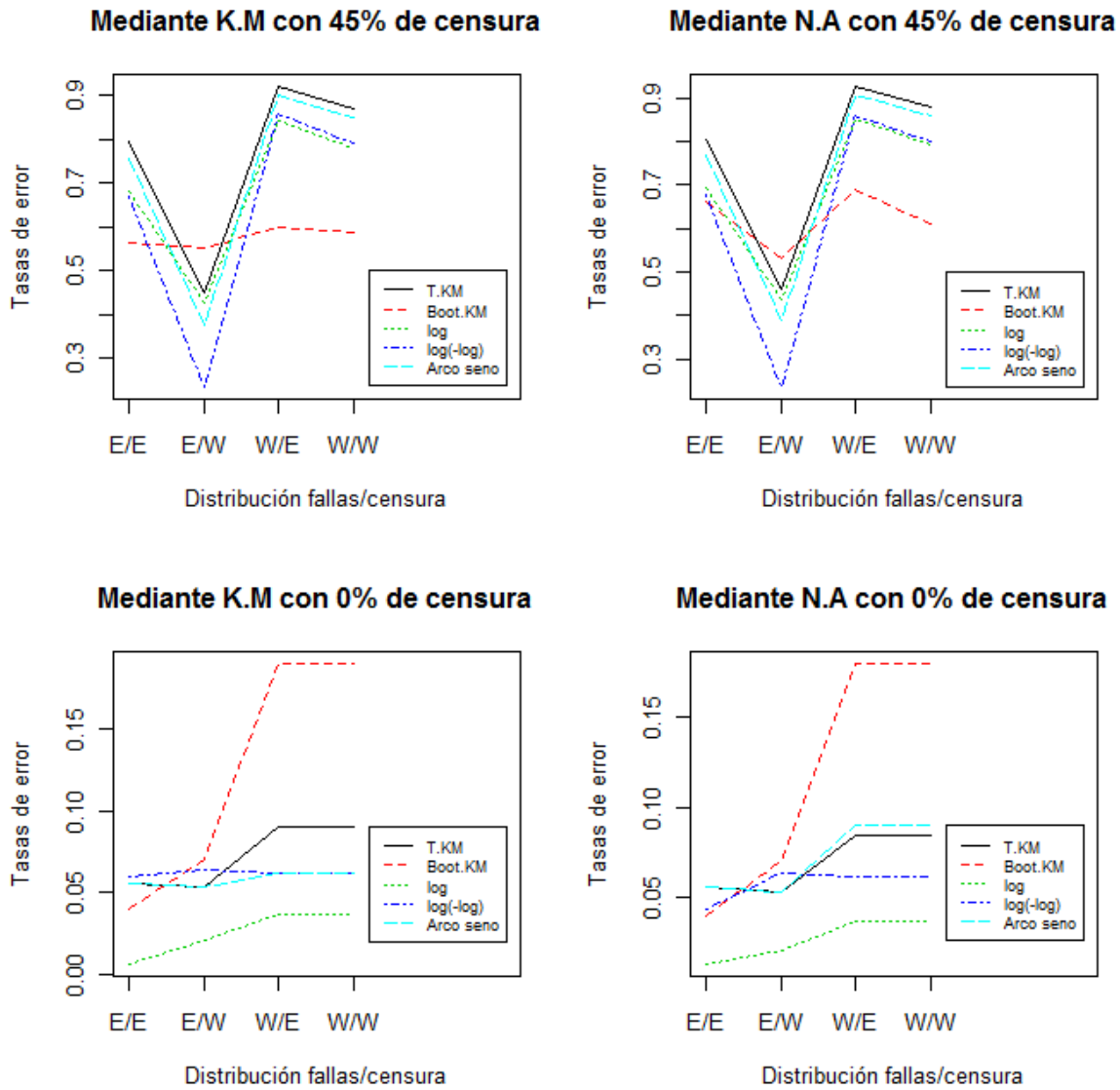


Figura 5.8: Tasas de Error de I.C para $S(t)$ con $n = 75$ y $n = 100$

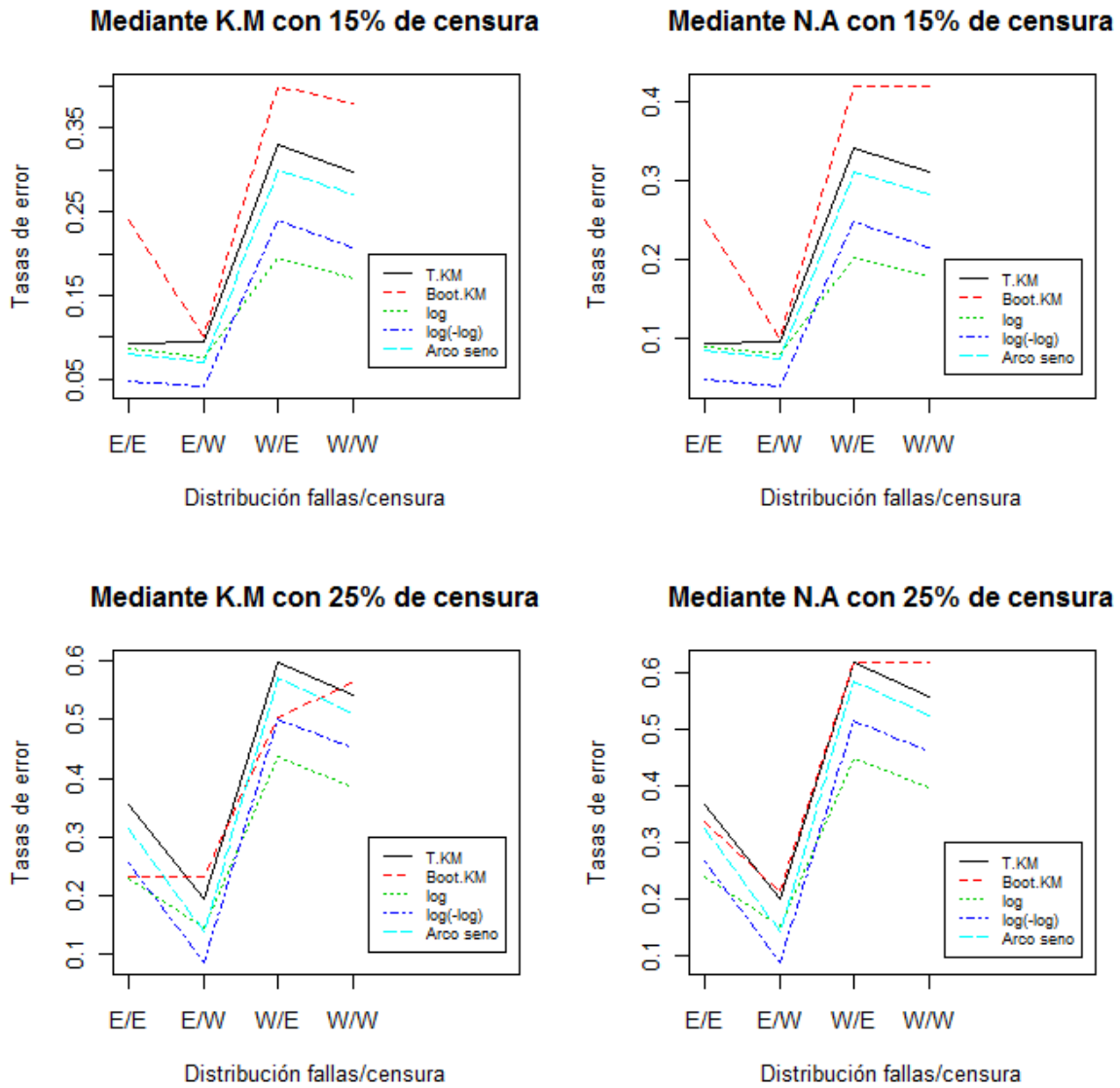


Figura 5.9: Tasas de Error de I.C para $S(t)$ con $n = 100$

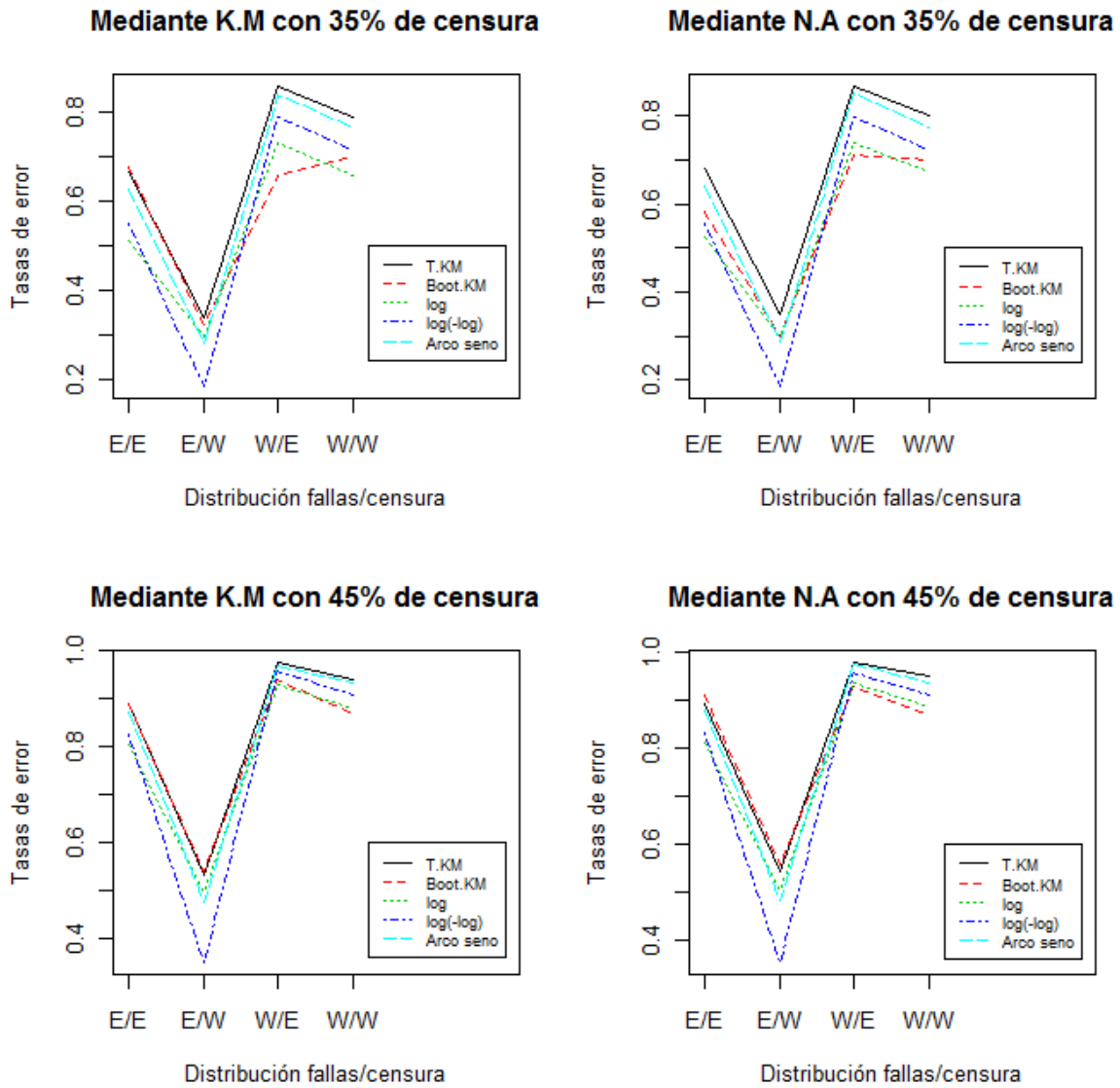


Figura 5.10: Tasas de Error de I.C para $S(t)$ con $n = 100$

5.2. Documentación Funciones y Códigos en R

`gen.datos`

Descripción

Genera los datos de supervivencia para los tiempos de falla y los tiempos de censura mediante las distribuciones predeterminadas, para este trabajo se utilizaron combinaciones de Exponencial y Weibull.

Uso

```
gen.datos(n,pc,lambda1,lambda2)
```

Argumentos

n: Tamaño de la muestra

lambda1: Parámetro de la distribución

lambda2: Parámetro de la distribución

IC.median

Descripción

Calcula los intervalos de confianza que se compararan como son los tradicionales, mediante las transformaciones \log , $\log(-\log)$ y Arcsen

```
IC.median(datos,tmed)
```

Argumentos

datos: Datos generados mediante la función anterior

tmed: Tiempo de interés

S1

Descripción

Calcula la probabilidad de supervivencia del tiempo de interés ($t = 0.2$) mediante el estimador de Kaplan-Meier, según los datos generados mediante (gen.datos) descrito anteriormente.

Uso

S1(datos)

Argumentos

datos: Datos generados

S2**Descripción**

Calcula la probabilidad de supervivencia del tiempo de interés ($t = 0.2$) mediante el estimador de Nelson-Aalen, según los datos generados mediante (gen.datos) descrito anteriormente.

Uso

S2(datos)

Argumentos

datos: Datos generados

boot.ci**Descripción**

Calcula los intervalos de confianza bootstrap mediante de los estimadores de Kaplan-Meier y Nelson-Aalen.

Uso

```
boot.ci(remuestra1,type="norm")
```

Argumentos

remuestra1: Muestra tomada usando el comando `boot(datos, S1, R = 500, sim = "ordinary")`, donde, `datos` y `S1` son descritos anteriormente, `R` corresponde al número de replicas, y `sim` representa el tipo de simulación ordinaria, tomando los de tipo normal.

5.2.1. Código

Esta función genera $n-pc*n$ tiempos de falla y $pc*n$ tiempos de censura. Los tiempos de falla se generan de una $\text{exp}(\lambda_1)$ y los tiempos de censura de una $\text{exp}(\lambda_2)$

```
gen.datos=function(n,pc,lambd1,lambd2){
  n.event=round((1-pc)*n)
  n.cen=n-n.event
  t.event=matrix(rexp(n.event,lambd1),ncol=1)
  t.cen=matrix(rexp(n.cen,lambd2),ncol=1)
  time=rbind(t.event,t.cen)
  status=rbind(matrix(rep(1,n.event),ncol=1),matrix(rep(0,n.cen),ncol=1))
  datos=data.frame(time,status)
  return(datos)
}
```

Esta función estima los IC a comparar.

```
alpha=0.05
IC.median=function(datos,tmed){
  a=survfit(with(datos, Surv(time,status)))
  pos.t=which.min(abs(a$time-tmed))\
  if (a$time[pos.t]<tmed){j=pos.t}else{
  j=pos.t+1}
  S.tmed=a$surv[j]
  S2.tmed=exp(-sum((a$n.event[1:j]/a$n.risk[1:j])))
  varianza de greenwood
```

```

vari=S.tmed^2*sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))
varianza de Aalen-Johansen\\
vari2=S2.tmed^2*sum((a$.event[1:j]/(a$.risk[1:j])^{2}))
LC tradicionales K-M
LI=S.tmed-qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)*S.tmed
LS=S.tmed+qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)*S.tmed
IC=c(LI,LS)
LC tradicionales N-A
LIA=S2.tmed-qnorm(1-alpha/2)*sqrt(vari2)
LSA=S2.tmed+qnorm(1-alpha/2)*sqrt(vari2)
ICA=c(LIA,LSA)
LC transformación log -K-M
LI2=S.tmed*exp(-qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)/S.tmed)
LS2=S.tmed*exp(qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)/S.tmed)
IC2=c(LI2,LS2)
LC transformación log -N-A
LIA2=S2.tmed*exp(-qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/
(a$.risk[1:j])^{2}))))/S2.tmed)
LSA2=S2.tmed*exp(qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/
(a$.risk[1:j])^{2}))))/S2.tmed)
ICA2=c(LIA2,LSA2)
LC transformación log(-log)-K-M
LI3=S.tmed^exp(-qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)/(log(S.tmed)))
LS3=S.tmed^exp(+qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)/(log(S.tmed)))
IC3=c(LI3,LS3)
LC transformación log(-log)-N-A
LIA3=S2.tmed^exp(-qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)/(log(S2.tmed)))
LSA3=S2.tmed^exp(+qnorm(1-alpha/2)*sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)/(log(S2.tmed)))
ICA3=c(LIA3,LSA3)
LC transformación Arco seno -K-M
LI4=(sin(max(0,asin(sqrt(S.tmed))-0.5*qnorm(1-alpha/2)*
sqrt(sum((a$.event[1:j]/a$.risk[1:j])/
(a$.risk[1:j]*(1-a$.event[1:j]/a$.risk[1:j])))*)*
sqrt(S.tmed/(1-S.tmed))))))^{2}

```

```

LS4=(sin(min(pi/2,asin(sqrt(S.tmed))+0.5*qnorm(1-alpha/2)*
sqrt(sum((a$n.event[1:j]/a$n.risk[1:j])/(a$n.risk[1:j]*
(1-a$n.event[1:j]/a$n.risk[1:j]))))*sqrt(S.tmed/(1-S.tmed))))))^2}
IC4=c(LI4,LS4)
LC transformación Arco seno -N-A
LIA4=(sin(max(0,asin(sqrt(S2.tmed))-0.5*qnorm(1-alpha/2)*
sqrt(sum((a$n.event[1:j]/
(a$n.risk[1:j])^2)))*sqrt(S2.tmed/(1-S2.tmed))))))^2}
LSA4=(sin(min(pi/2,asin(sqrt(S2.tmed))+0.5*qnorm(1-alpha/2)*
sqrt(sum((a$n.event[1:j]/
(a$n.risk[1:j])^2)))*sqrt(S2.tmed/(1-S2.tmed))))))^2}
ICA4=c(LIA4,LSA4)

list("IC"=IC,"IC2"=IC2,"IC3"=IC3,"IC4"=IC4,
"ICA"=ICA,"ICA2"=ICA2,"ICA3"=ICA3,"ICA4"=ICA4)}

```

Esta parte hace la simulación de las muestras y calcula los IC

```

NS<-2000
IC=matrix(0,ncol=2,nrow=NS)
IC2=matrix(0,ncol=2,nrow=NS)
IC3=matrix(0,ncol=2,nrow=NS)
IC4=matrix(0,ncol=2,nrow=NS)
ICA=matrix(0,ncol=2,nrow=NS)
ICA2=matrix(0,ncol=2,nrow=NS)
ICA3=matrix(0,ncol=2,nrow=NS)
ICA4=matrix(0,ncol=2,nrow=NS)
ICA5=matrix(0,ncol=2,nrow=NS)
for (i in 1:NS){
set.seed(i)
datos=gen.datos(50,0.15,2,1)
tiempo.mediano=0.2
alpha=0.05
IC[i,]=IC.median(datos,tiempo.mediano)$IC
IC2[i,]=IC.median(datos,tiempo.mediano)$IC2
IC3[i,]=IC.median(datos,tiempo.mediano)$IC3
IC4[i,]=IC.median(datos,tiempo.mediano)$IC4
ICA[i,]=IC.median(datos,tiempo.mediano)$ICA
ICA2[i,]=IC.median(datos,tiempo.mediano)$ICA2
ICA3[i,]=IC.median(datos,tiempo.mediano)$ICA3
ICA4[i,]=IC.median(datos,tiempo.mediano)$ICA4
IIC=cbind(IC,IC2,IC3,IC4,ICA,ICA2,ICA3,ICA4)
}

```

```
colnames(IIC)<-c("LI(K-M)", "LS(K-M)", "LI(K-M) (log)", "LS(K-M) (log)",
"LI(K-M) (log(-log))", "LS(K-M) (log(-log))", "LI(K-M) (Arcsen)",
"LS(K-M) (Arcsen)", "LI(N-A)", "LS(N-A)", "LI(N-A) (log)", "LS(N-A) (log)",
"LI(N-A) (log(-log))", "LS(N-A) (log(-log))", "LI(N-A) (Arcsen)",
"LS(N-A) (Arcsen)")
```

Calcular tasas de error, y el índice en los IC

```
param=exp(-tiempo.mediano*2)
contador=(ifelse(IIC[,1]<=param & IIC[,2]>=param,1,0))
contador2=(ifelse(IIC[,3]<=param & IIC[,4]>=param,1,0))
contador3=(ifelse(IIC[,5]<=param & IIC[,6]>=param,1,0))
contador4=(ifelse(IIC[,7]<=param & IIC[,8]>=param,1,0))
contadorA=(ifelse(IIC[,9]<=param & IIC[,10]>=param,1,0))
contadorA2=(ifelse(IIC[,11]<=param & IIC[,12]>=param,1,0))
contadorA3=(ifelse(IIC[,13]<=param & IIC[,14]>=param,1,0))
contadorA4=(ifelse(IIC[,15]<=param & IIC[,16]>=param,1,0))

t.e=sum(contador,na.rm=T)/(length(contador)-sum(is.na(contador)))
t.e2=sum(contador2,na.rm=T)/(length(contador2)-sum(is.na(contador2)))
t.e3=sum(contador3,na.rm=T)/(length(contador3)-sum(is.na(contador3)))
t.e4=sum(contador4,na.rm=T)/(length(contador4)-sum(is.na(contador4)))
t.eA=sum(contadorA,na.rm=T)/(length(contadorA)-sum(is.na(contadorA)))
t.eA2=sum(contadorA2,na.rm=T)/(length(contadorA2)-sum(is.na(contadorA2)))
t.eA3=sum(contadorA3,na.rm=T)/(length(contadorA3)-sum(is.na(contadorA3)))
t.eA4=sum(contadorA4,na.rm=T)/(length(contadorA4)-sum(is.na(contadorA4)))

t.e=cbind(1-t.e,1-t.e2,1-t.e3,1-t.e4,1-t.eA,1-t.eA2,1-t.eA3,1-t.eA4)
NR=1-t.e
NN=1-alpha

LI1=LI2=LI3=LI4=LI5=LI6=LI7=LI8=numeric(0)
for (i in 1:NS){
LI1[i]=IIC[i,2]-IIC[i,1]
LI2[i]=IIC[i,4]-IIC[i,3]
LI3[i]=IIC[i,6]-IIC[i,5]
LI4[i]=IIC[i,8]-IIC[i,7]
LI5[i]=IIC[i,10]-IIC[i,9]
LI6[i]=IIC[i,12]-IIC[i,11]
LI7[i]=IIC[i,14]-IIC[i,13]
LI8[i]=IIC[i,16]-IIC[i,15]
```

```
}

```

```
LPI=cbind(mean(LI1,na.rm=T),mean(LI2,na.rm=T),mean(LI3,na.rm=T),
mean(LI4,na.rm=T),mean(LI5,na.rm=T),mean(LI6,na.rm=T)
,mean(LI7,na.rm=T),mean(LI8,na.rm=T))

```

```
IN=(2-mean(LPI))/2*NR/NN

```

Realiza la simulación y calcula los I.C \$bootstrap\$

```
NS=2000
IC1=matrix(0,ncol=2,nrow=NS)
IC2=matrix(0,ncol=2,nrow=NS)
for (k in 1:NS){
  set.seed(k)
  datos=gen.datos(75,0.15,1,0.5)
  remuestra1<-boot(datos,S1,R=300,sim="ordinary")
  remuestra2<-boot(datos,S2,R=300,sim="ordinary")
  IC1[k,]=boot.ci(remuestra1,type="norm")\$normal[2:3]
  IC2[k,]=boot.ci(remuestra2,type="norm")\$normal[2:3]
  IIC=cbind(IC1,IC2)
}

```

Calcular tasas de error en los IC

```
param=exp(-0.2*2)
contador=array(0,dim=NS)
contador2=array(0,dim=NS)
for (i in 1:NS){
  if (param<=IIC[i,2] & param>=IIC[i,1]){
    contador[i]=1} else{
    contador[i]=0}
  if (param<=IIC[i,4] & param>=IIC[i,3]){
    contador2[i]=1} else{
    contador2[i]=0}
}
t.e=1-sum(contador)/length(contador)
t.e2=1-sum(contador2)/length(contador2)

```

tasas de error de todos los IC

```
t.e=cbind(t.e,t.e2)

```



```
NR=1-t.e
NN=0.95
LI1=LI2=numeric(0)
for (i in 1:NS){
LI1[i]=IIC[i,2]-IIC[i,1]
LI2[i]=IIC[i,4]-IIC[i,3]
}

LPI=cbind(mean(LI1),mean(LI2))
IN=(2-mean(LPI))/2*NR
```