

Estimación de Kaplan Meier Bootstrap de la Curva de Supervivencia

Freddy Tineo G.¹, Ysela Agüero P.²,

Emma Cambillo M.²

Resumen

En este trabajo se estima la función de supervivencia mediante el método no paramétrico conocido como el estimador de Kaplan Meier Bootstrap, bajo el supuesto de normalidad asintótica.

Palabras Clave: *Análisis de supervivencia. Tasas de riesgo. Estimador de Kaplan Meier Bootstrap.*

Abstract

In this work the function of survival is estimated by means of the non-parametric method known like the Kaplan Meier Bootstrap estimator, under the assumption of asymptotical normality.

Keywords: *Analysis of survival. Rates of risk. Estimator of Kaplan Meier Bootstrap.*

1. Introducción

El análisis estadístico de los datos referidos a tiempos de vida, de falla o de respuesta se ha convertido en un tema de creciente interés por sus aplicaciones en diversas áreas como medicina, ingeniería, y las ciencias biológicas. Esta metodología de análisis de datos se aplica en diversas investigaciones desde aquellas referidas a resistencia de artículos manufacturados hasta investigaciones sobre enfermedades del ser humano.

En el área biomédica, es muy común que la variable que se desea estudiar sea el tiempo que tarda en ocurrir un evento, ya sea beneficioso (ej. curación, alta hospitalaria) o perjudicial (muerte, aparición de efecto adverso). La metodología para analizar este tipo de datos

¹Ministerio de Agricultura, Lima-Perú

²UNMSM, Facultad de Ciencias Matemáticas, Lima - Perú.

se conoce en medicina como "Análisis de Supervivencia", debido a que fueron diseñados inicialmente para estudiar el tiempo de supervivencia de pacientes, fundamentalmente en el campo de la oncología.

La característica más importante de los datos en este tipo de investigaciones es que, al final del periodo de observación no todos los pacientes presentan el suceso objeto de estudio. Además, puede ocurrir que algunos individuos hayan abandonado el estudio por causas diversas, no siendo posible determinar su estado al final del estudio. Cuando el suceso es la muerte, el fallecimiento puede ocurrir por causas diferentes a la del estudio.

Otra característica de este tipo de estudios es que, los pacientes pueden ser incorporados en él durante todo el periodo de observación, de modo que los últimos en hacerlo serán observados un periodo menor, comparado con los que entraron al inicio del estudio. Así la probabilidad de que les ocurra el evento de interés será menor y al final del estudio pueden haber pacientes que aún no presentan el suceso. Estas observaciones incompletas son denominadas censuradas ("censored").

Uno de los objetivos del análisis de supervivencia es estimar la función de supervivencia, las tasas de riesgo, el tiempo medio de supervivencia, y otras características, para lo cual se pueden utilizar métodos paramétricos y no paramétricos.

Un método no paramétrico muy utilizado para estimar la función de supervivencia es el estimador de Kaplan Meier o límite del producto. La inferencia estadística utilizando este estimador permite construir intervalos de confianza para la función de supervivencia en tiempos fijados t , bajo el supuesto de normalidad asintótica.

Pero, no siempre es posible disponer de tamaños de muestra suficientemente grandes para garantizar la normalidad asintótica del estimador de Kaplan Meier, por lo que se podría utilizar el método de remuestreo "bootstrap" propuesto por Efron (1981) para aproximar la distribución de este estimador.

El objetivo de este trabajo es obtener estimaciones de Kaplan Meier Bootstrap de la función de supervivencia de una muestra de pacientes adultos de sexo femenino con diagnóstico de Linfoma No Hodgkin. El estudio fue realizado en un importante Instituto de Investigación en enfermedades oncológicas del Perú.

En la sección 2 se define el estimador de Kaplan Meier o límite del producto, sus propiedades y el procedimiento de cálculo.

En la sección 3 se trata brevemente sobre el método de remuestreo bootstrap, los pasos básicos para la construcción de la distribución bootstrap, la descripción de los principales métodos para calcular los intervalos de confianza y el método de remuestreo bootstrap aplicado a la estimación mediante la estadística de Kaplan Meier.

Finalmente, en la sección 4 se presentan los resultados de la aplicación del método de remuestreo bootstrap para la estimación de las bandas de confianza de los tiempos de supervivencia de pacientes con diagnóstico de Linfoma No Hodgkin. Los datos proceden de una muestra de pacientes cuyas edades varían entre 40 y 84 años de edad. Los resultados son comparados con las estimaciones de Kaplan Meier asintóticas.

2. Estimación de Kaplan Meier para la curva de supervivencia

En esta sección describiremos brevemente el método de estimación no paramétrico de Kaplan Meier para la función de supervivencia, el cual permite analizar los tiempos de vida observados calculando la función de supervivencia empírica, percentiles y otras estadísticas útiles.

Si la muestra observada no contiene datos censurados, la función de supervivencia empírica (FSE) es definida como:

$$\hat{S}(t) = \frac{\text{Número de observaciones} \geq t}{n} ; \quad t \geq 0 \quad (2.1)$$

La cual es una función escalonada que decrece en $1/n$ justo después de cada observación del tiempo de vida de un individuo, siempre que todas las observaciones sean distintas. Más generalmente, si hay d tiempos de vida iguales a t , la función de supervivencia empírica (FSE) decrece en una cantidad d/n inmediatamente después de t .

Cuando se tiene datos censurados, se requiere realizar algunas modificaciones en (2.1), puesto que generalmente, no se conocerá exactamente el número de tiempos de vida mayores o iguales a t . Kaplan-Meier (1958) propusieron una modificación de (2.1) a la cual denominaron "estimador del límite del producto" (LP) de la función de supervivencia. El estimador se define como sigue.

Definición

Supóngase que se dispone de las observaciones de los tiempos de supervivencia de n individuos y que hay k ($k \leq n$) tiempos distintos en los cuales el evento de interés -digamos la muerte- ocurre, esto es, $t_1 < t_2 < \dots < t_k$. Admitiendo la posibilidad de tener más de una muerte en t_j ; d_j representará el número de muertes en t_j . Además, existen los tiempos de censura L_i para individuos para quienes el evento no fue observado. El estimador del límite del producto, $\hat{S}(t)$, es definido como:

$$P(T > t) = \hat{S}(t) = \prod_{j:t_j < t} \frac{n_j - d_j}{n_j} \quad (2.2)$$

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 . Si un tiempo de censura L_i y un tiempo de vida t_j son registrados como iguales, se adoptará la convención de que el tiempo de censura es ajustado aumentando una cantidad infinitesimal a la derecha; para que L_i sea considerado infinitesimalmente mayor que t_j . En otras palabras, cualquier individuo con tiempo de censura igual a t_j será incluido en el conjunto de n_j individuos en riesgo en t_j . Esta convención es razonable, puesto que un individuo censurado en el tiempo L casi ciertamente sobrevive después de L .

Otro punto sobre (2.2) involucra situaciones en las que el mayor tiempo observado en la muestra sea un tiempo de censura. En este caso, la estimación del límite del producto se toma como definido sólo hasta esta última observación.

2.1. El estimador del límite del producto como un *Estimador de Máxima Verosimilitud (EMV)*

El estimador (2.2) posee un número importante de propiedades, el principal es que $\hat{S}(t)$ es un estimador consistente de $S(t)$ bajo condiciones generales.

Supongamos que en el intervalo $[t_{j-1}, t_j)$ se han observado λ_j tiempos censurados L_i^j ($i = 1, \dots, \lambda_j$). Haciendo $t_0 = 0$, $t_{k+1} = \infty$, y $j = 1, 2, \dots, k+1$, se incluye tiempos de censura que ocurren antes de la primera o después de la última observación de tiempos de vida. La función de supervivencia teórica para los 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 continua por la izquierda. Bajo las suposiciones con respecto a los tipos de censuras (censura por la derecha), la función de verosimilitud observada es de la forma (Lawless, 1982),

$$L = \prod_{j=1}^k \left[\left(\prod_{i=1}^{\lambda_j} S(L_i^j) \right) [S(t_j) - S(t_j + 0)]^{d_j} \right] \prod_{i=1}^{\lambda_{k+1}} S(L_i^{k+1}). \quad (2.3)$$

Utilizando métodos iterativos se obtiene la estimación de la función de supervivencia que maximiza la función de verosimilitud (2.3).

2.2. Intervalo de Confianza del estimador de Kaplan Meier

Para evaluar eficazmente los resultados; cuando se usa el estimador del límite del producto; es conveniente tener un estimador de la varianza de $\hat{S}(t)$, el cual es dado por,

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

Las propiedades asintóticas de los estimadores de máxima verosimilitud garantizan la normalidad asintótica del estimador de Kaplan Meier. Luego, los intervalos de $(1 - \alpha) \%$ de confianza de la función de supervivencia en cada tiempo fijo t pueden ser calculados de la siguiente manera.

$$\left\langle \hat{S}(t_j) - Z_{\alpha/2} \sqrt{\hat{V}ar [\hat{S}(t_j)]} < S(t_j) < \hat{S}(t_j) + Z_{\alpha/2} \sqrt{\hat{V}ar [\hat{S}(t_j)]} \right\rangle$$

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

3. Método de remuestreo Bootstrap

El primero en considerar la técnica del bootstrap de una manera sistemática fue Efron (1979). El nombre "bootstrapping" es usado por Efron para seguir las ideas de Tukey (1958) quien usó el nombre de "Jackknifing" para describir su método de remuestreo. Se supone que

la técnica de Tukey es análoga al uso de “la navaja de bolsillo” de un explorador, y que el método de Efron es análoga a “alguien saliendo de un pantano, jalándose los pasadores de sus propias botas”.

La esencia del “bootstrapping” es la idea que, en ausencia de cualquier otro conocimiento sobre la distribución de una población, la distribución de los valores observados en una muestra aleatoria de tamaño n de la población es la mejor información para aproximar su distribución de probabilidades.

El bootstrap, ha recibido mucha atención y es muy popular, su importancia radica en que constituye un método para estimar la distribución de probabilidades de una estadística a partir de muestras finitas y como tal, puede ser usado para construir intervalos confidenciales no paramétricos. Efron (1982), demostró mediante estudios de simulación que estos intervalos de confianza son más exactos que aquellos basados en la distribución asintótica del estimador.

En situaciones simples, la incertidumbre asociada a un estimador; puede ser resuelta mediante cálculos analíticos suponiendo un determinado modelo probabilístico asociado a las variables en cuestión, Cochran (1977). Sin embargo, en situaciones más complejas, la solución analítica puede ser complicada y sus resultados sesgados si se realizan simplificaciones o supuestos inapropiados. En estas condiciones, los métodos bootstrap, Efron y Tibishirani (1996) proporcionan una alternativa válida ya que no asumen que la muestra haya sido tomada de una población con una distribución probabilística concreta. La idea básica del bootstrap es que en ausencia de otra información sobre la población, la distribución de una muestra aleatoria es la mejor guía para determinar la distribución de la población; por lo tanto, remuestrear con reemplazamiento la muestra aleatoria nos aproxima a lo que sucedería si se remuestrea la población, Manly (1997).

Debido a que el sustento teórico matemático-estadístico del bootstrap es bastante complejo, hasta finales de la década del '80 del siglo pasado, la eficiencia del método era verificada sólo de manera empírica, es decir, en el terreno de la práctica, Fernández (1999).

El método bootstrap es robusto debido a que no se asume ninguna distribución particular de la población y ha sido aplicado a diversos problemas de estimación, sustituyendo la complejidad de las derivaciones analíticas por el poder de cálculo de los ordenadores.

Para seleccionar las muestras bootstrap se obtiene un número de réplicas suficientemente grande de muestras de tamaño n , obtenidas con reemplazamiento, siendo n el número de observaciones de la muestra original.

El estimador bootstrap no aporta nada que no pueda solucionarse mediante los métodos analíticos clásicos, sin embargo, para estimadores complejos donde el cálculo analítico del error estándar no es evidente, el remuestreo bootstrap proporciona un método alternativo para calcularlo.

3.1. Intervalos de Confianza Bootstrap

Existen muchos métodos para calcular intervalos de confianza de un parámetro θ mediante el bootstrap, los más frecuentemente usados son los siguientes.

1. Límites de Confianza Bootstrap estándar

El principio aquí es que, si un estimador $\hat{\theta}$ es normalmente distribuido con media θ y desviación estándar σ existe entonces una probabilidad de $1 - \alpha$ que el intervalo

$$\theta - Z_{\alpha/2}\sigma < \hat{\theta} < \theta + Z_{\alpha/2}\sigma$$

contenga cualquier valor aleatorio de $\hat{\theta}$, donde $Z_{\alpha/2}$ es el valor que se excede con probabilidad $\alpha/2$ para una distribución normal estándar. Esta afirmación es equivalente a:

$$\hat{\theta} - Z_{\alpha/2}\sigma < \theta < \hat{\theta} + Z_{\alpha/2}\sigma. \quad (3.1)$$

Utilizando la expresión (3.1), se obtiene el intervalo confidencial bootstrap estándar para θ a partir de

$$\text{Estimador} \pm Z_{\alpha/2}\sigma_{BOOT}$$

Donde,

$$\sigma_{\theta}^* = \sqrt{\frac{\sum_{b=1}^B (\theta_b^* - \bar{\theta})^2}{(B-1)}} = \sigma_{BOOT} \quad (3.2)$$

donde $\bar{\theta}$ es la media de las réplicas bootstrap, que corresponde al estimador bootstrap del estadístico θ .

$$\bar{\theta} = \frac{1}{B} \sum_{b=1}^B \theta_b^*$$

De (3.1), el intervalo confidencial bootstrap estándar para θ es estimado calculando la desviación estándar de las B estimaciones bootstrap $\hat{\theta}_j$ ($j=1,2,\dots,B$), donde B es el número de muestras bootstrap. El análisis de la distribución de frecuencias bootstrap (distribución bootstrap de $\hat{\theta}$) permite observar hasta que punto se puede asumir la normalidad.

Cuando se observa que no es posible asumir la normalidad del estimador se pueden estimar los límites de confianza utilizando otros métodos basados en la propia distribución de frecuencias. El siguiente es uno de estos métodos.

2. Límites de Confianza mediante el método del percentil

Varias aproximaciones para construir los límites de confianza que han recibido considerable atención; se basan en tentativas de aproximar los percentiles de la distribución de

un estimador utilizando los percentiles generados por el "bootstrapping". Efron (1979) lo describió como el *método del percentil*.

En este método los límites del $100(1 - \alpha)\%$ de confianza para un parámetro son precisamente los dos valores que contienen el $100(1 - \alpha)\%$ central de las B estimaciones bootstrapping obtenidas desde la muestra original. Esto se justifica en la suposición que existe una transformación que permite aproximar la distribución del estimador a una normal.

El remuestreo bootstrap de la muestra de tamaño n se utiliza para generar la distribución bootstrap del estimador del parámetro de interés. El intervalo $100(1 - \alpha)\%$ de confianza para el parámetro es dado por dos valores que incluyen el $100(1 - \alpha)\%$ central de esta distribución. Por ejemplo, un intervalo de confianza del 95% es dado por el valor que excede $2,5\%$ y $97,5\%$ del área de la distribución generada.

Existen otros métodos para calcular intervalos de confianza mediante el bootstrap tales como: el método de percentil con corrección del sesgo, el bootstrap-t, BC_α (Intervalo de confianza corregido por sesgo acelerado) y el método ABC (Intervalo de confianza aproximado por bootstrap), entre otros (DiCiccio y Efron, 1996).

3.2. El método de remuestreo bootstrap aplicado a los 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 por ejemplo, una gran cantidad de datos censurados, reduciendo el número de observaciones con información completa.

Efron ([3]), sugirió utilizar el método bootstrap para estimar la supervivencia utilizando datos de registros diarios de muertes y el estimador de Kaplan Meier. Un tratamiento amplio del método bootstrap aplicado al análisis de supervivencia y una discusión sobre su utilidad para la construcción de bandas de confianza para funciones de supervivencia, pueden encontrarse en Reid (1981) y especialmente en Efron (1981) y Akritas (1986).

3.2.1. Estimación bootstrap para datos censurados

Consideremos el caso de datos censurados por la derecha, esto es de la forma dada por $\{(x_1, d_1), (x_2, d_2), \dots, (x_n, d_n)\}$, donde x_j es la j -ésima observación, censurada o no, y d_j es una función indicadora,

$$d_j = \begin{cases} 1 & \text{si } x_j \text{ no censurado} \\ 0 & \text{si } x_j \text{ censurado} \end{cases}$$

El estimador bootstrap $\hat{\sigma}_{BOOT}$ presentado en (2.3) es apropiado e igual en este caso, sólo que los datos ahora son los pares (x_j, d_j) , $j = 1, 2, \dots, n$.

El procedimiento de estimación es el siguiente:

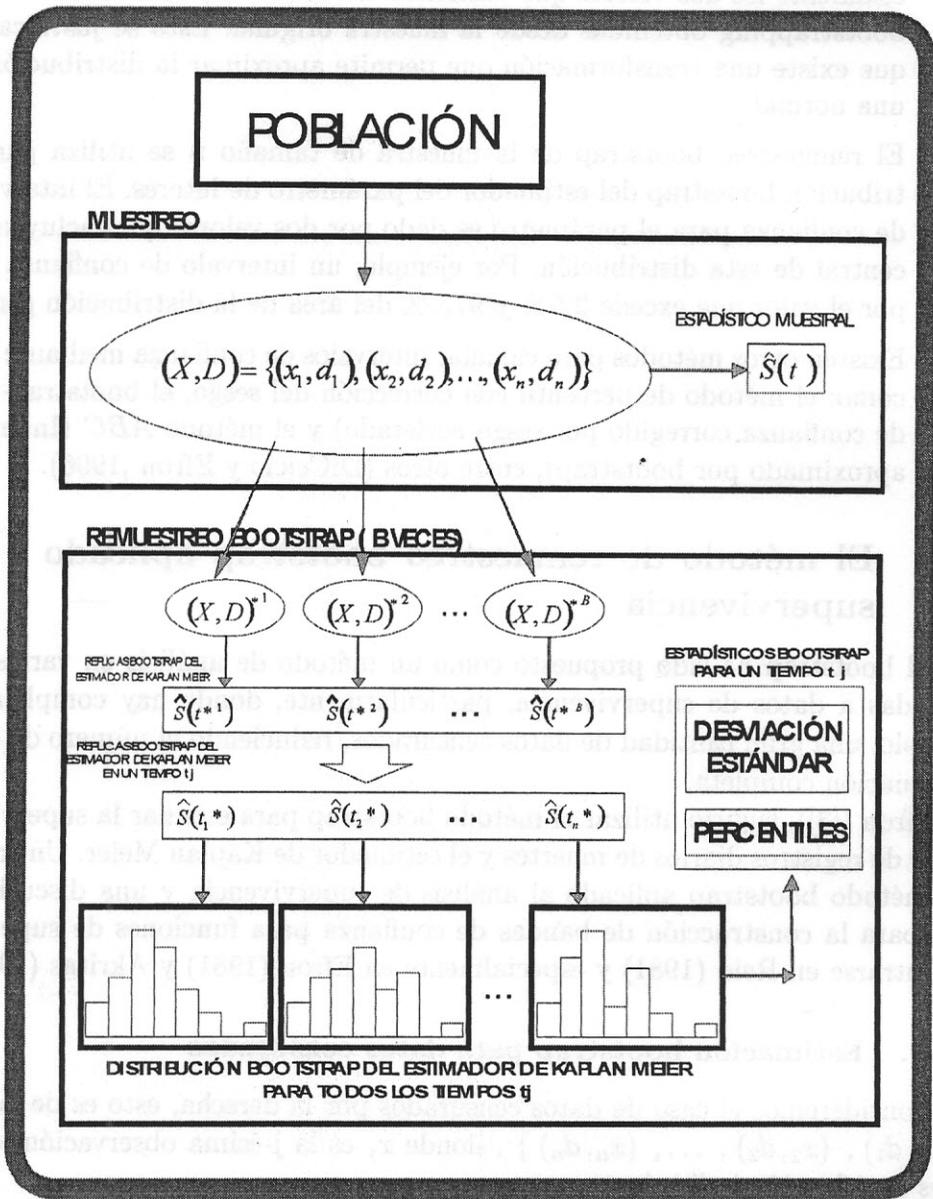


Figura 1: Procedimiento de estimación de bandas de confianza mediante el método Kaplan Meier Bootstrap

1. Selecciona una muestra bootstrap $(X_1^*, D_1^*), (X_2^*, D_2^*), \dots, (X_n^*, D_n^*)$ muestreando independientemente n con reemplazamiento desde \hat{F} , la distribución que pone masa $1/n$ en cada punto (x_j, d_j)
2. Denotar con **datos*** el conjunto de datos artificiales, y calcular $S(t_{j*}) = S(\text{datos*})$
3. Repetir independientemente los pasos (1) y (2) B veces, hasta obtener el conjunto de estimaciones $S(t_j^{*1}), S(t_j^{*2}), \dots, S(t_j^{*B})$.
4. Calcular $\hat{\sigma}_{BOOT}$ como en (3.2).

Este procedimiento bootstrap sólo requiere que los pares observados (x_j, d_j) de todas las submuestras sean observaciones independientes e idénticamente distribuidas (iid) desde una distribución F en $\mathcal{R}^1 \times \{0, 1\}$.

En la figura N° 1 se muestra el proceso de estimación de Kaplan Meier bootstrap con datos censurados (desviaciones estándar y las bandas de confianza). El algoritmo para el cálculo de los intervalos de confianza de las probabilidades de supervivencia por el método de Kaplan Meier Bootstrap se encuentra en Tineo (2005).

4. Aplicación

Frecuentemente, los oncólogos necesitan determinar el tiempo medio de supervivencia, o las probabilidades de supervivencia de pacientes atacados con diferentes tipos de cáncer por lo que emprenden investigaciones mediante estudios de cohortes de pacientes con diagnóstico positivo de la enfermedad.

El problema que surge a menudo es que usualmente, las muestras son pequeñas y además durante el periodo fijado para la observación se pierden pacientes, algunos de ellos abandonan el tratamiento y otros mueren por otras causas. Esto hace que, las muestras se reduzcan, por lo tanto, se necesita utilizar métodos de estimación que sean robustos cuando las muestras son pequeñas y que además tomen en cuenta los datos censurados.

En esta aplicación se utilizarán los datos de un estudio referido a una cohorte de pacientes con cierto tipo de cáncer denominado Linfoma no Hodgkin, el cual describiremos brevemente.

Los linfomas son tumores que se originan en los ganglios linfáticos, y que pueden tener un origen extra ganglionar como en la médula hematopoyética el tubo digestivo, pulmón, piel, bazo, o dondequiera que se encuentre tejido linfoide asociado a las mucosas. Estos linfomas se clasifican en: linfoma de Hodgkin o enfermedad de Hodgkin y todos los otros tipos que se agrupan en los linfomas No Hodgkin.

Métodos y Procedimientos

Los datos corresponden a un estudio de intervención realizado por un importante Instituto de Investigación de enfermedades oncológicas; durante el periodo 1990 - 2000. En esta investigación se observaron los tiempos de supervivencia de una muestra de 100 pacientes

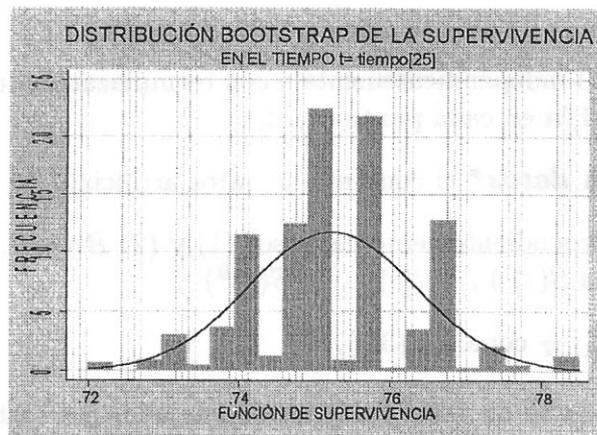


Figura 2: Distribución Bootstrap del estimador de Kaplan Meier para $t=3.47$ meses

de sexo femenino con edades entre 40 y 84 años quienes fueron diagnosticados con Linfoma No Hodgkin en los estadios II y III. El 37 por ciento de los pacientes de la muestra sobrevivió hasta el final del estudio (censura de tipo I).

A partir de estos datos se realizó la estimación de las bandas de confianza para las curvas de supervivencia mediante los métodos de Kaplan Meier para toda la muestra y Kaplan Meier bootstrap. Luego se procedió a comparar las estimaciones obtenidas por ambos métodos.

La desviación estándar para el intervalo de confianza Kaplan Meier para la muestra total se estimó mediante la fórmula de Greenwood, y las respectivas bandas de confianza se obtuvieron bajo el supuesto de normalidad asintótica. Por otro lado, las bandas de confianza bootstrap fueron obtenidas mediante el método de percentiles, considerando un total de 670 muestras bootstrap (B). Las simulaciones se realizaron utilizando un programa elaborado en el paquete estadístico STATA versión 8.2. Los datos y la sintaxis del programa computacional se encuentran en Tineo ((2005), Anexo A).

Resultados

Para cumplir con los objetivos planteados se obtuvieron estimaciones de las probabilidades de supervivencia por ambos métodos y luego se procedió a la comparación de la eficiencia de las estimaciones obtenidas.

Para el cálculo de los intervalos de confianza de la curva de supervivencia mediante el remuestreo bootstrap, utilizando el método del percentil, se obtuvieron 670 replicaciones (muestras bootstrap), luego se construyó la distribución bootstrap del estimador de Kaplan Meier para cada tiempo t_i ($i = 1, 2, \dots, 100$), generándose 100 distribuciones bootstrap. Luego, se calcularon los percentiles 2,5 y 97,5, para cada distribución, obteniéndose de esta manera los intervalos de confianza al 95% de la probabilidad de supervivencia, para cada t_i . A continuación se muestra la distribución bootstrap para la observación $t=3,47$.

Se puede observar que para la observación $t=3,47$ la distribución de las probabilidades de supervivencia no es normal, esto mismo ocurre cuando se calcula las distribuciones para

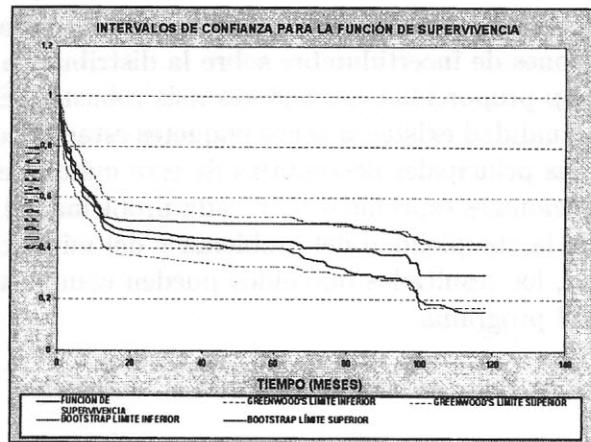


Figura 3: Bandas de confianza estimadas para la función de Supervivencia.

la mayoría de las observaciones, lo cual muestra que el método bootstrap es una buena alternativa para estimar la curva de supervivencia, sin hacer ninguna suposición con respecto a la distribución del estimador.

La tabla N° 1 muestra los tiempos de supervivencia observados, las probabilidades de supervivencia y los intervalos de confianza por ambos métodos.

La figura N° 3 muestra las bandas de confianza para la curva de supervivencia de los pacientes con infoma no Hodking, calculadas mediante el método de Kaplan Meier (fórmula de Greenwood) y mediante el método de Kaplan Meier bootstrap (método del percentil), utilizando 670 replicaciones bootstrap (muestras bootstrap). Notar que los intervalos de confianza bootstrap son más estrechos. En los últimos periodos de tiempo el número de observaciones es reducido por lo que ambos intervalos son igualmente ineficientes.

5. Conclusiones y Recomendaciones

En este artículo se ha tratado de mostrar de manera práctica que el método de remuestreo bootstrap aplicado al estimador de Kaplan Meier proporciona intervalos de confianza de la curva de supervivencia más eficientes, comparado con los intervalos de confianza de Kaplan Meier, obtenidos a partir de la muestra total bajo el supuesto de normalidad asintótica del estimador.

El método de remuestreo bootstrap aplicado al estimador de Kaplan Meier es muy útil para el análisis de supervivencia cuando los tamaños de muestra son pequeños y fuertemente censurados, no teniendo importancia el tipo de censura que presenta la muestra analizada.

El método de remuestreo bootstrap puede fallar debido a la presencia de casos atípicos, puesto que el bootstrap asume que la distribución de la muestra representa la distribución de la población, si la muestra es inusual, también lo serán los estimadores bootstrap derivados de ella. Tampoco existe ningún método estadístico que pueda resolver problemas derivados de tamaños de muestra muy pequeños o malos diseños de muestreo. En estas circunstancias,

el método de remuestreo bootstrap puede fallar como fallaría cualquier otro método. Sin embargo, en situaciones de incertidumbre sobre la distribución del estimador, el método de remuestreo bootstrap proporciona estimadores más robustos. Efron y Tibshirani (1993).

Aunque en la actualidad existen muchos paquetes estadísticos para realizar el remuestreo bootstrap, una de las principales desventajas de este método es la necesidad de desarrollar programas computacionales específicos para cada problema. El tiempo que se emplea en los cálculos depende de la complejidad del problema y del número de réplicas (muestras bootstrap). Sin embargo, los resultados obtenidos pueden compensar todo el esfuerzo realizado en la elaboración del programa.

Referencias

- [1] AKRITAS, M.G. Bootstrapping the Kaplan Meier Estimator. *Journal of the American Statistical Association*, 81: 1032-1038. 1986.
- [2] DICICCIO, T.J. y EFRON B., Bootstrap Confidence Intervals. *Statistic Science*, 11: 189-228. 1996.
- [3] EFRON, B. Bootstrap methods: Another look at jackknife. *The Annals of Statistics*, 7: 1-26. 1979.
- [4] EFRON, B. Censored Data and the Bootstrap. *Journal of the American Statistical Association*, 76: 312-319. 1981.
- [5] EFRON, B. *The Jackknife, the Bootstrap, and other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics. 1982.
- [6] EFRON B y TIBSHIRANI . *An introduction to the bootstrap*. Chapman & Hall. New York. EEUU. 1993
- [7] FERNANDEZ, J.A. El método Jacknife para la Estimación de Varianzas de Estimadores. Universidad Nacional Mayor de San Marcos, FCM. 1999.
- [8] LAWLESS, J.F. *Statistical Models and Methods for Lifetime Data*. University of Waterloo. 1982.
- [9] MANLY, B.F.J. *Randomization Bootstrap and Monte Carlo Methods in Biology*, 2º ed. London. Chapman y Hall. 1997.
- [10] TINEO, F. Estimación de Kaplan Meier bootstrap de la Curva de Supervivencia. Monografía. Universidad Nacional Mayor de San Marcos, FCM. 2005.

TABLA N° 1 : INTERVALOS DE CONFIANZA PARA LA FUNCIÓN DE SUPERVIVENCIA
 ESTIMADO POR LA FÓRMULA DE GREENWOOD Y POR EL MÉTODO DEL PERCENTIL BOOTSTRAP
 B = 670 REPLICACIONES BOOTSTRAP

TIEMPO (MESES)	FUNCIÓN DE SUPERVIVENCIA ESTIMADA	GREENWOOD		BOOTSTRAP	
		LÍMITE INFERIOR	LÍMITE SUPERIOR	LÍMITE INFERIOR	LÍMITE SUPERIOR
0.10	0.98	0.953	1.000	0.960	0.990
0.10	0.98	0.953	1.000	0.960	0.990
0.20	0.97	0.937	1.000	0.940	0.980
0.37	0.96	0.922	0.998	0.930	0.970
0.40	0.94	0.893	0.987	0.920	0.970
0.40	0.94	0.893	0.987	0.920	0.970
0.47	0.94	0.893	0.987	0.909	0.950
0.70	0.93	0.880	0.980	0.900	0.949
0.73	0.92	0.866	0.973	0.880	0.940
0.77	0.91	0.853	0.966	0.870	0.930
0.90	0.90	0.841	0.959	0.869	0.919
1.03	0.89	0.828	0.951	0.859	0.909
1.17	0.88	0.815	0.943	0.849	0.898
1.27	0.87	0.803	0.936	0.839	0.888
1.57	0.86	0.791	0.928	0.830	0.878
1.60	0.85	0.779	0.919	0.810	0.868
1.70	0.83	0.755	0.903	0.800	0.857
1.70	0.83	0.755	0.903	0.800	0.857
1.93	0.82	0.743	0.895	0.790	0.837
2.10	0.81	0.731	0.886	0.778	0.826
:	:	:	:	:	:
:	:	:	:	:	:
90.63	0.37	0.275	0.472	0.293	0.460
90.80	0.37	0.275	0.472	0.293	0.460
91.00	0.37	0.275	0.472	0.281	0.460
92.23	0.37	0.275	0.472	0.281	0.457
94.77	0.37	0.275	0.472	0.278	0.457
95.10	0.37	0.275	0.472	0.269	0.457
96.57	0.37	0.275	0.472	0.263	0.450
97.73	0.34	0.239	0.450	0.260	0.450
97.97	0.34	0.239	0.450	0.258	0.445
98.47	0.34	0.239	0.450	0.249	0.439
98.90	0.34	0.239	0.450	0.239	0.438
99.60	0.34	0.239	0.450	0.230	0.438
99.63	0.34	0.239	0.450	0.207	0.436
100.37	0.30	0.168	0.422	0.198	0.428
102.37	0.30	0.168	0.422	0.181	0.422
105.93	0.30	0.168	0.422	0.174	0.422
108.27	0.30	0.168	0.422	0.169	0.422
111.13	0.30	0.168	0.422	0.165	0.422
116.10	0.30	0.168	0.422	0.153	0.422
118.13	0.30	0.168	0.422	0.153	0.422