

Tópicos de análisis de sobrevivida: segunda parte

Gabriel Cavada Ch.^{1,2}

¹Facultad de Medicina, Universidad de los Andes.

²División de Bioestadística, Escuela de Salud Pública, Universidad de Chile.

Topics survival analysis: part two

Estimación de la Función de Sobrevivida

Estimador de Kaplan-Meier

Un problema fundamental que se debe abordar es estimar la función de sobrevivida cuando existen censuras en la muestra. Para resolver este problema se han propuesto algunas alternativas tales como: el método actuarial, el método de las tablas de vida y el método de Efron; sin embargo la mejor propuesta que se ha hecho es la conocida estimación de Kaplan y Meier, que fue publicada en el JASA (Journal American Statistics Association) el año 1958, que se expondrá con detalle a continuación.

La estimación de Kaplan-Meier es un método no paramétrico (no asume ninguna función de probabilidad para la función de sobrevivida) y por máxima verosimilitud, es decir se basa en maximizar la función de verosimilitud (es decir la probabilidad de haber observado una determinada colección de datos) de la muestra (metodología que se conoce como verosimilitud generalizada). Una muestra aleatoria de tamaño n , extraída de una población, estará formada por k ($k \leq n$) tiempos $t_1 < t_2 < \dots < t_k$ en los que se observan eventos. En cada tiempo t_i existen n_i "individuos en riesgo de morir" (elementos de la muestra para los que el evento puede ocurrir, o que $T \geq t_i$) y se observan d_i eventos (muertes). Además en el intervalo $[t_i, t_{i+1})$ se producen m_i pérdidas (censuras).

La función de verosimilitud para toda la muestra es:

$$L = \prod_{i=1}^k \binom{n_i}{d_i} h_i^{d_i} (1 - h_i)^{n_i - d_i} \quad (1)$$

pues se asume que en cada intervalo de tiempo los eventos siguen una distribución binomial con $n = n_i$ y $p = h_i$. Para construir esta función se ha asumido que la información

contenida en las pérdidas es que, para cada una de ellas, el evento ocurre en un tiempo mayor que el tiempo en que se observa la pérdida. Maximizando esta función se encuentra que el estimador de h_i en cada intervalo es:

$$\hat{h}_i = \frac{d_i}{n_i} \quad \text{con } i = 1, 2, \dots, k \quad (2)$$

Es decir la cantidad de muertes observadas en un intervalo dividida por la cantidad de sujetos expuestos a riesgo de morir al inicio del intervalo

Recordando que la función de sobrevivida, para tiempos discretos se puede escribir como:

$$S(t_k) = \prod_{i=1}^{k-1} (1 - h(t_i)) \quad (3)$$

Pues la probabilidad de sobrevivir a un determinado intervalo de tiempo es una probabilidad condicionada a haber llegado vivo a dicho intervalo. Así al reemplazar h por su estimador (ecuación 2) en la ecuación 3 se obtiene el estimador de la función de sobrevivida de Kaplan-Meier (también denominada estimador producto límite):

$$\hat{S}(t_i) = \prod_{i=1}^{k-1} \left(1 - \frac{d_i}{n_i}\right) \quad \text{con } i = 1, 2, \dots, k$$

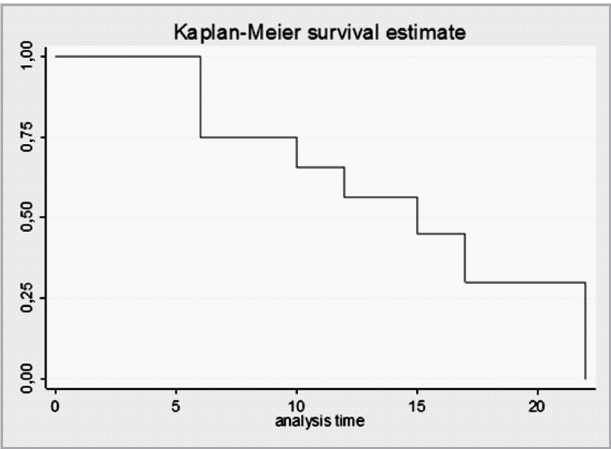
Ejemplo:

Se sigue en el tiempo a 12 individuos con una prótesis cardíaca y se encuentran los siguientes tiempos de sobrevivida en años: 6+, 6, 6, 6, 10, 12+, 12, 15, 15+, 17, 22, 22, donde el signo más indica censura; es decir se perdieron 3 individuos en los tiempos 6, 12 y 15. La manera más cómoda de calcular los estimadores anteriores es disponer los datos en una tabla como la siguiente:

Rincón de la Bioestadística

La función de sobrevivida estimada se puede representar gráficamente como sigue:

Tiempo (t _i)	Eventos (d _i)	Censuras	Sujetos en riesgo (n _i)	Función de riesgo (h _i)	1-h _i	Función de sobrevivida S(t _i)
0	0	0	12	0	1	1
6	3	1	8	$\frac{3}{12}$	$\frac{9}{12}$	$\frac{9}{12}=0,75$
10	1	0	7	$\frac{1}{8}$	$\frac{7}{8}$	$\frac{9}{12} \cdot \frac{7}{8}=0,66$
12	1	1	5	$\frac{1}{7}$	$\frac{6}{7}$	$\frac{9}{12} \cdot \frac{7}{8} \cdot \frac{6}{7}=0,56$
15	1	1	3	$\frac{1}{5}$	$\frac{4}{5}$	$\frac{9}{12} \cdot \frac{7}{8} \cdot \frac{6}{7} \cdot \frac{4}{5}=0,45$
17	1	0	2	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{9}{12} \cdot \frac{7}{8} \cdot \frac{6}{7} \cdot \frac{4}{5} \cdot \frac{2}{3}=0,30$
22	2	0	0	1	0	0



El gráfico de la estimación de sobrevivida de Kaplan-Meier es muy informativo, pues el ancho del escalón, indica la magnitud del intervalo de tiempo en que se observaron muertes, mientras que el alto indica la magnitud de las incidencias de muerte en dicho intervalo.