Construção do Estimador de Kaplan-Meier

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida
- Censura à direita

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida
- Censura à direita
- Amostra: $Y_i = \min(T_i, C_i)$ e $\delta_i = \mathbf{I}(T_i < C_i)$, $i = 1, ..., n$

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida
- Censura à direita
- Amostra: $Y_i = \min(T_i, C_i)$ e $\delta_i = \mathbf{I}(T_i < C_i)$, $i = 1, ..., n$
- com $r \leq n$ tempos distintos, $t_1 < t_2 < \cdots < t_r$

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida
- Censura à direita
- Amostra: $Y_i = \min(T_i, C_i)$ e $\delta_i = \mathbf{I}(T_i < C_i)$, $i = 1, ..., n$
- com $r \leq n$ tempos distintos, $t_1 < t_2 < \cdots < t_r$
- com d_i falhas no tempo t_i

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida
- Censura à direita
- Amostra: $Y_i = \min(T_i, C_i)$ e $\delta_i = \mathbf{I}(T_i < C_i)$, $i = 1, ..., n$
- com $r \leq n$ tempos distintos, $t_1 < t_2 < \cdots < t_r$
- com d_i falhas no tempo t_i
- e n_i indivíduos em risco no tempo t_i^-

Construção do estimador de Kaplan-Meier

- T: tempo até ocorrência de um evento, origem bem definida
- Censura à direita
- Amostra: $Y_i = \min(T_i, C_i)$ e $\delta_i = \mathbf{I}(T_i < C_i)$, $i = 1, ..., n$
- com $r \leq n$ tempos distintos, $t_1 < t_2 < \cdots < t_r$
- com d_i falhas no tempo t_i
- e n_i indivíduos em risco no tempo t_i^-
- Encontrar um estimador para $S(t) = P(T > t)$

Paralelo com Tábua de Vida

- Particionar o tempo em $\xi_0, \xi_1, ..., \xi_p$

$$S(\xi_2)$$

Paralelo com Tábua de Vida

- Particionar o tempo em $\xi_0, \xi_1, ..., \xi_p$

$$S(\xi_2) = P(T > \xi_2)$$

Paralelo com Tábua de Vida

- Particionar o tempo em $\xi_0, \xi_1, ..., \xi_p$

$$\begin{aligned} S(\xi_2) &= P(T > \xi_2) \\ &= P(T > \xi_2 \mid T > \xi_1) P(T > \xi_1) \end{aligned}$$

Paralelo com Tábua de Vida

- Particionar o tempo em $\xi_0, \xi_1, ..., \xi_p$

$$
\begin{aligned}
S(\xi_2) &= P(T > \xi_2) \\
&= P(T > \xi_2 \mid T > \xi_1) P(T > \xi_1) \\
&= P(T > \xi_2 \mid T > \xi_1) P(T > \xi_1 \mid T > \xi_0) \cdot 1
\end{aligned}
$$

Paralelo com Tábua de Vida

- Particionar o tempo em $\xi_0, \xi_1, ..., \xi_p$

$$\begin{aligned} S(\xi_2) &= P(T > \xi_2) \\ &= P(T > \xi_2 \mid T > \xi_1)P(T > \xi_1) \\ &= P(T > \xi_2 \mid T > \xi_1)P(T > \xi_1 \mid T > \xi_0) \cdot 1 \end{aligned}$$

- Seja $p_i = P(T > \xi_i \mid T > \xi_{i-1}),\ i = 1, ..., r$ e $t_0 = 0$

Paralelo com Tábua de Vida

- Particionar o tempo em $\xi_0, \xi_1, ..., \xi_p$

$$
\begin{aligned}
S(\xi_2) &= P(T > \xi_2) \\
&= P(T > \xi_2 \mid T > \xi_1) P(T > \xi_1) \\
&= P(T > \xi_2 \mid T > \xi_1) P(T > \xi_1 \mid T > \xi_0) \cdot 1
\end{aligned}
$$

- Seja $p_i = P(T > \xi_i \mid T > \xi_{i-1}),\ i = 1, ..., r$ e $t_0 = 0$
- Podemos escrever

$$
S(\xi_i) = p_i p_{i-1} p_{i-2} \ldots p_1 = \prod_{k=1}^{i} p_k
$$

Paralelo com Tábua de Vida

- Seja $q_i = 1 - p_i = P(T < \xi_i \mid T > \xi_{i-1})$

Paralelo com Tábua de Vida

- Seja $q_i = 1 - p_i = P(T < \xi_i \mid T > \xi_{i-1})$
- Estimativa para q_i

$$\hat{q}_i = \hat{P}(T \leq \xi_i \mid T > \xi_{i-1})$$

Paralelo com Tábua de Vida

- Seja $q_i = 1 - p_i = P(T < \xi_i \mid T > \xi_{i-1})$
- Estimativa para q_i

$$\begin{aligned} \hat{q}_i &= \hat{P}(T \leq \xi_i \mid T > \xi_{i-1}) \\ &= \frac{\text{eventos em } (\xi_{i-1}, \xi_i]}{\text{indivíduos vivos até } \xi_{i-1}} \end{aligned}$$

Paralelo com Tábua de Vida

- Seja $q_i = 1 - p_i = P(T < \xi_i \mid T > \xi_{i-1})$
- Estimativa para q_i

$$
\begin{aligned}
\hat{q}_i &= \hat{P}(T \leq \xi_i \mid T > \xi_{i-1}) \\
&= \frac{\text{eventos em } (\xi_{i-1}, \xi_i]}{\text{indivíduos vivos até } \xi_{i-1}} \\
&= \frac{d_i}{n_i - w_i/2}
\end{aligned}
$$

Paralelo com Tábua de Vida

- Seja $q_i = 1 - p_i = P(T < \xi_i \mid T > \xi_{i-1})$
- Estimativa para q_i

$$\begin{aligned}\hat{q}_i &= \hat{P}(T \leq \xi_i \mid T > \xi_{i-1}) \\ &= \frac{\text{eventos em } (\xi_{i-1}, \xi_i]}{\text{indivíduos vivos até } \xi_{i-1}} \\ &= \frac{d_i}{n_i - w_i/2}\end{aligned}$$

- Estimativa para p_i

$$\Rightarrow \hat{p}_i = 1 - \frac{d_i}{n_i - w_i/2}$$

$$\hat{S}(\xi_i) = \prod_{j: t_j \leq \xi_i} \hat{p}_j = \prod_{j: t_j \leq \xi_i} \left[1 - \frac{d_j}{n_j - w_j/2}\right], \ \hat{S}(\xi_0) = 1$$

Estimador de Kaplan-Meier

Estimador de Kaplan-Meier

$$\hat{S}(t) = \begin{cases} 1 & \text{se } t < t_1 \\ \prod_{t_i \leq t} \left[1 - \frac{d_i}{n_i}\right] & \text{se } t \geq t_1 \end{cases}$$

Exemplo

Estudo com pacientes com câncer de mama (Liu, 2012), que foram acompanhadas por um período de 7 anos, até o óbito ou fim do estudo:

$$5, 17, 20+, 24, 32, 35+, 40, 46, 47, 50, 59, 74+$$

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	$0/12$	$12/12$	1.00

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	$0/12$	$12/12$	1.00
1	5	1	0	12	$1/12$	$11/12$	0.92

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	0/12	12/12	1.00
1	5	1	0	12	1/12	11/12	0.92
2	17	1	0	11	1/11	10/11	0.83

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	$0/12$	$12/12$	1.00
1	5	1	0	12	$1/12$	$11/12$	0.92
2	17	1	0	11	$1/11$	$10/11$	0.83
3	24	1	1	9	$1/9$	$8/9$	0.74

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	0/12	12/12	1.00
1	5	1	0	12	1/12	11/12	0.92
2	17	1	0	11	1/11	10/11	0.83
3	24	1	1	9	1/9	8/9	0.74
4	32	1	0	8	1/8	7/8	0.65

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	$0/12$	$12/12$	1.00
1	5	1	0	12	$1/12$	$11/12$	0.92
2	17	1	0	11	$1/11$	$10/11$	0.83
3	24	1	1	9	$1/9$	$8/9$	0.74
4	32	1	0	8	$1/8$	$7/8$	0.65
5	40	1	1	6	$1/6$	$5/6$	0.54

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	$0/12$	$12/12$	1.00
1	5	1	0	12	$1/12$	$11/12$	0.92
2	17	1	0	11	$1/11$	$10/11$	0.83
3	24	1	1	9	$1/9$	$8/9$	0.74
4	32	1	0	8	$1/8$	$7/8$	0.65
5	40	1	1	6	$1/6$	$5/6$	0.54
6	46	1	0	5	$1/5$	$4/5$	0.43

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	0/12	12/12	1.00
1	5	1	0	12	1/12	11/12	0.92
2	17	1	0	11	1/11	10/11	0.83
3	24	1	1	9	1/9	8/9	0.74
4	32	1	0	8	1/8	7/8	0.65
5	40	1	1	6	1/6	5/6	0.54
6	46	1	0	5	1/5	4/5	0.43
7	47	1	0	4	1/4	3/4	0.32

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	0/12	12/12	1.00
1	5	1	0	12	1/12	11/12	0.92
2	17	1	0	11	1/11	10/11	0.83
3	24	1	1	9	1/9	8/9	0.74
4	32	1	0	8	1/8	7/8	0.65
5	40	1	1	6	1/6	5/6	0.54
6	46	1	0	5	1/5	4/5	0.43
7	47	1	0	4	1/4	3/4	0.32
8	50	1	0	3	1/3	2/3	0.22

Exemplo

j	t_j	d_j	w_j	n_j	$\hat{q}_j$	$\hat{p}_j$	$\hat{S}(t_j)$
0	0	0	0	12	0/12	12/12	1.00
1	5	1	0	12	1/12	11/12	0.92
2	17	1	0	11	1/11	10/11	0.83
3	24	1	1	9	1/9	8/9	0.74
4	32	1	0	8	1/8	7/8	0.65
5	40	1	1	6	1/6	5/6	0.54
6	46	1	0	5	1/5	4/5	0.43
7	47	1	0	4	1/4	3/4	0.32
8	50	1	0	3	1/3	2/3	0.22
9	59	1	0	2	1/2	1/2	0.11

Variância

Formula de Greenwood

$$\widehat{\text{Var}}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i\,(n_i - d_i)}$$

Intervalo de confiança

$$\frac{\hat{S}(t) - S(t)}{\sqrt{\widehat{\mathrm{Var}}[\hat{S}(t)]}} \rightarrow \mathrm{N}(0,1)$$

$$IC(S(t);\gamma) = \left[\hat{S}(t) - z_{\gamma/2}\sqrt{\widehat{\mathrm{Var}}[\hat{S}(t)]}; \quad \hat{S}(t) + z_{\gamma/2}\sqrt{\widehat{Var}[\hat{S}(t)]}\right]$$

Intervalo de confiança

$$\frac{\hat{S}(t) - S(t)}{\sqrt{\widehat{\text{Var}}[\hat{S}(t)]}} \to \text{N}(0,1)$$

$$IC(S(t);\gamma) = \left[\hat{S}(t) - z_{\gamma/2}\sqrt{\widehat{\text{Var}}[\hat{S}(t)]}; \quad \hat{S}(t) + z_{\gamma/2}\sqrt{\widehat{Var}[\hat{S}(t)]}\right]$$

Intervalo de confiança

- Transformações
  - Kalbfleisch-Prentice

$$\hat{U}(t) = \log(-\log \hat{S}(t))$$

$$\widehat{\mathrm{Var}}(\hat{U}(t)) = \frac{1}{(\log \hat{S}(t))^2} \sum_{j:t_{(j)}<t} q_i / \left(n_j \left(1 - q_i\right)\right)$$

$$IC(S(t);\gamma) = \left[(\hat{S}(t))^{-\theta}; (\hat{S}(t))^{\theta}\right]$$

com $\theta = \exp\left\{z_{\gamma/2}\sqrt{\widehat{Var}[\hat{U}(t)]}\right\}$