



RELATORIO:

**APLICAÇÃO DO ESTIMADOR DE KAPLAN-MEIER E ESTIMADOR DE NELSON-
AALEN**

Lucas Manoel Batista de Albuquerque

Análise de Sobrevida

2023

1 INTRODUÇÃO

A análise de sobrevivência é uma técnica estatística amplamente utilizada em diversas áreas, como medicina, engenharia, biologia, entre outras. Ela tem como objetivo estudar o tempo até a ocorrência de um evento de interesse, como o tempo até a falha de um equipamento, o tempo de vida de um paciente ou o tempo até a ocorrência de um evento de mercado. Com o uso da análise de sobrevivência, é possível estimar a probabilidade de um evento ocorrer em um determinado período de tempo e identificar os fatores que podem influenciar nesse tempo. Neste artigo, serão apresentados dois exemplos de aplicação da análise de sobrevivência em diferentes áreas, ilustrando como essa técnica pode ser útil na tomada de decisões e na compreensão de fenômenos complexos.

2 DADOS USADOS COMO EXEMPLO

Exemplo 1: Um estudo clínico aleatorizado foi realizado para investigar o efeito da terapia com esteroide no tratamento de hepatite viral aguda (Gregory et al., 1976). Vinte e nove pacientes com esta doença foram aleatorizados para receber um placebo ou o tratamento com esteroide. Cada paciente foi acompanhado por 16 semanas ou até a morte (evento de interesse) ou até a perda de acompanhamento. Os tempos de sobrevivência observados, em semanas, para os dois grupos são apresentados na Tabela 1. O símbolo + indica censura.

Tabela 1 - Tempos, em semanas, observados no estudo de hepatite

Grupos	Tempos de Sobrevivência
Controle	1+, 2+, 3, 3, 3+, 5+, 5+, 16+, 16+, 16+, 16+, 16+, 16+, 16+, 16+
Esteróide	1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+

Exemplo 2: Um estudo experimental realizado com camundongos para verificar a eficácia da imunização pela malária foi conduzido no Centro de Pesquisas René Rachou, Fiocruz, MG. Nesse estudo, quarenta e quatro camundongos foram aleatoriamente divididos em três grupos e todos foram infectados pela malária (*Plasmodium berguei*). Os camundongos do grupo 1 foram imunizados 30 dias antes da infecção. Além da infecção pela malária, os camundongos dos grupos 1 e 3 foram, também, infectados pela esquistossomose (*Schistosoma mansoni*). A resposta de interesse nesse estudo foi o tempo decorrido desde a infecção pela malária até a morte do camundongo. Este tempo foi medido em dias e o estudo foi acompanhado por 30 dias. Os tempos de sobrevivência observados para os três grupos encontram-se na Tabela 2. O símbolo + indica censura.

Tabela 2 - Tempos, em dias, observados no estudo de malária

Grupos	Tempos de Sobrevivência
Grupo 1 (16)	7, 8, 8, 8, 8, 12, 12, 17, 18, 22, 30+, 30+, 30+, 30+, 30+, 30+
Grupo 2 (15)	8, 8, 9, 10, 10, 14, 15, 15, 18, 19, 21, 22, 22, 23, 25
Grupo 3 (13)	8, 8, 8, 8, 8, 8, 9, 10, 10, 10, 11, 17, 19

3 METODOLOGIA

3.1 ESTIMADOR DE KAPLAN-MEIER

O estimador de Kaplan-Meier é uma técnica estatística não-paramétrica que é comumente utilizada para estimar a função de sobrevivência em estudos de sobrevivência. Esse estimador leva em conta dados censurados, ou seja, dados que não apresentam informações completas sobre a duração do evento de interesse. O estimador de Kaplan-Meier leva em conta esses dados censurados e é capaz de produzir uma curva de sobrevivência que mostra a proporção de indivíduos sobreviventes em diferentes momentos de tempo.

Construção do Estimador de Kaplan-Meier:

- Ordenar os tempos distintos de falha:

$$t_1 < t_2 < \dots < t_k$$

- Utilizando a seguinte notação:
 - d_i : número de falhas no tempo t_i ;
 - n_i : número de observações sob risco (não falhou e não foi censurado) até o tempo t_i (exclusivo);

O estimador de Kaplan-Meier é:

$$\hat{S}(t) = \prod_{\substack{i \\ t_j < t}} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{\substack{i \\ t_j < t}} \left(1 - \frac{d_i}{n_i} \right)$$

É definida como a probabilidade de uma observação não falhar até certo instante t , ou seja, a probabilidade desta sobreviver ao tempo t .

3.2 ESTIMADOR DE NELSON-AALEN

O estimador de Nelson-Aalen é um estimador não-paramétrico da função de acumulação de risco em estudos de sobrevivência. Esse estimador é amplamente utilizado em análise de sobrevivência para estimar a taxa de falha cumulativa ao longo do tempo, levando em conta dados censurados.

Uma forma de expressar a sua função de sobrevivência é:

$$S(t) = \exp(-\Lambda(t))$$

Onde: $\tilde{\Lambda}(t) = \sum_{t:t_i < t} \left(\frac{d_i}{n_i}\right)$

4 RESULTADOS

4.1 KAPLAN-MEIER

Para o Exemplo 1 temos:

Tabela 3 - Exemplo 1 - Estimador de Kaplan-Meier para o grupo controle

Tempo (t_i)	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
0	[0, 3)	15	0	0	1
3	[3, 16]	13	2	0,153	0,846

Tabela 4 - Exemplo 1 - Estimador Kaplan-Meier para o grupo que fez uso de esteroides

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
0	[0, 1)	14	0	0	1
1	[1, 5)	14	3	0,214	0,786
5	[5, 7)	9	1	0,11	0,698
7	[7, 8)	8	1	0,125	0,611
8	[8, 10)	7	1	0,142	0,524
10	[10, 16]	6	1	0,166	0,437

Para o Exemplo 2 temos:

Tabela 5 - Exemplo 2 – Estimador Kaplan-Meier para o Grupo 1

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
0	[0, 7)	16	0	0	1
7	[7, 8)	16	1	0,0625	0,938
8	[8, 12)	15	4	0,266	0,688
12	[12, 17)	11	2	0,1818	0,562
17	[17, 18)	9	1	0,111	0,5
18	[18, 22)	8	1	0,125	0,438
22	[22, +Inf)	7	1	0,1428	0,375

Tabela 6 - Exemplo 2 - Estimador Kaplan-Meier para o Grupo 2

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
8	[8, 9)	15	2	0,13	0,866
9	[9, 10)	13	1	0,076	0,8
10	[10, 14)	12	2	0,166	0,66
14	[14, 15)	10	1	0,1	0,6
15	[15, 18)	9	2	0,22	0,466
18	[18, 19)	7	1	0,1428	0,4
19	[19, 21)	6	1	0,166	0,33
21	[21, 22)	5	1	0,2	0,266
22	[22, 23)	4	2	0,5	0,133
23	[23, 25)	2	1	0,5	0,066
25	[25, +Inf)	1	1	1	0

Tabela 7 - Exemplo 2 - Estimador de Kaplan-Meier para o Grupo 3

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
8	[8, 9)	13	6	0,461	0,54
9	[9, 10)	7	1	0,143	0,46
10	[10, 11)	6	3	0,5	0,23
11	[11, 17)	3	1	0,33	0,15
17	[17, 19)	2	1	0,5	0,08
19	[19, +Inf)	1	1	1	0

4.2 NELSON-AALEN

Para o exemplo 1 temos:

Tabela 8 - Exemplo 1 - Estimador Nelson-Aalen para o grupo Controle

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
0	[0, 3)	15	0	0	1
3	[3, 16]	13	2	0,15	0,857

Tabela 9 - Exemplo 1 - Estimador Nelson-Aalen para o grupo que fez uso de esteroides

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
1	[1, 5)	14	3	0,214	0,807
5	[5, 7)	9	1	0,11	0,722
7	[7, 8)	8	1	0,125	0,637
8	[8, 10)	7	1	0,142	0,553
10	[10, +Inf)	6	1	0,166	0,468

Para o exemplo 2 temos:

Tabela 10 - Exemplo 2 - Estimador Nelson-Aalen para o Grupo 1

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
7	[7, 8)	16	1	0,0625	0,939
8	[8, 12)	15	4	0,266	0,72
12	[12, 17)	11	2	0,1818	0,6
17	[17, 18)	9	1	0,111	0,537
18	[18, 22)	8	1	0,125	0,474
22	[22, +Inf)	7	1	0,1428	0,411

Tabela 11 - Exemplo 2 - Estimador Nelson-Aalen para o Grupo 2

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
8	[8, 9)	15	2	0,13	0,875
9	[9, 10)	13	1	0,076	0,81
10	[10, 14)	12	2	0,166	0,686
14	[14, 15)	10	1	0,1	0,62
15	[15, 18)	9	2	0,22	0,497
18	[18, 19)	7	1	0,1428	0,43
19	[19, 21)	6	1	0,166	0,365
21	[21, 22)	5	1	0,2	0,3
22	[22, 23)	4	2	0,5	0,18
23	[23, 25)	2	1	0,5	0,11
25	[25, +Inf)	1	1	1	0,04

Tabela 12 - Exemplo 2 - Estimador Nelson-Aalen para o Grupo 3

t_i	Intervalo	n_i	d_i	$\frac{d_i}{n_i}$	$S(t)$
8	[8, 9)	13	6	0,461	0,63
9	[9, 10)	7	1	0,143	0,546
10	[10, 11)	6	3	0,5	0,331
11	[11, 17)	3	1	0,33	0,237
17	[17, 19)	2	1	0,5	0,144
19	[19, +Inf)	1	1	1	0,053

5 DISCUSSÃO

No exemplo 1 vemos que aparentemente há uma diferença significativa na quantidade ou frequência em que os pacientes morrem. O número de pacientes sob risco (n_i) do grupo controle foi maior ao final do estudo, indicando uma maioria sobrevivente. Além disso, no grupo controle as mortes ocorreram apenas nas 3 primeiras semanas, enquanto que no grupo que fez o uso de esteroides as mortes ocorreram de maneira quase uniforme ao longo do tempo. Também é possível observar esse comportamento pela probabilidade $S(t)$ que em ambos os estimadores decai gradualmente para o grupo que fez uso de esteroides.

No exemplo 2 é possível observar que apenas o grupo 1, ou seja, o que recebeu a imunização da malária é o que tem o maior número de sobreviventes até o final do estudo, evidenciado pelo número de pacientes sob risco (n_i). De maneira análoga, a probabilidade $S(t)$ é a que menos decai ao longo do tempo para o grupo 1, diferentemente do grupo 2 e 3 que ao final do estudo estão bem próximas de 0 para ambos os estimadores.

Nas Figuras 1, 2, 3 e 4 vemos as curvas de probabilidade de $S(t)$ para ambos os exemplos e estimadores.

Figura 1 - Curva de Probabilidade $S(t)$ do estimador de Kaplan-Meier para o exemplo 1. À esquerda representando o grupo 1 e à direita o grupo 2.

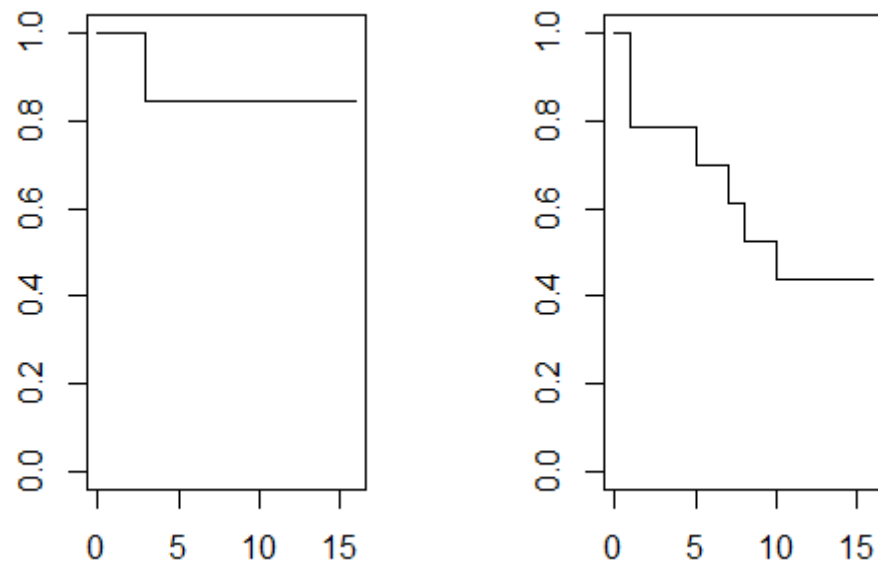


Figura 2 - Curva de Probabilidade $S(t)$ do estimador de Kaplan-Meier para o exemplo 2. Da esquerda para a direita está o grupo 1, grupo 2 e grupo 3.

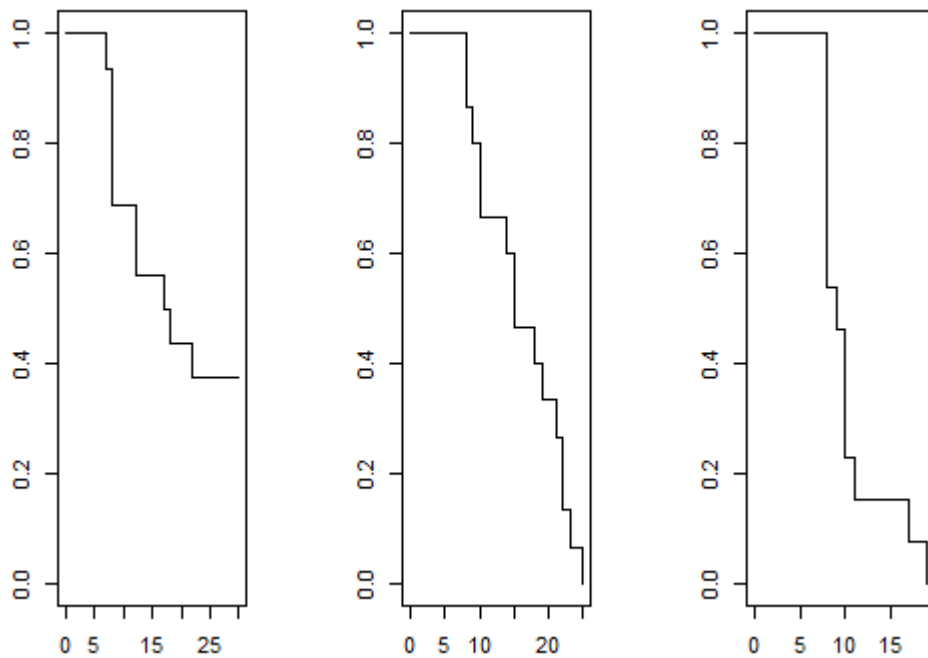


Figura 3 - Curva de Probabilidade $S(t)$ do estimador de Nelson-Aalen para o exemplo 1. Na esquerda o grupo 1 e na direita o grupo 2

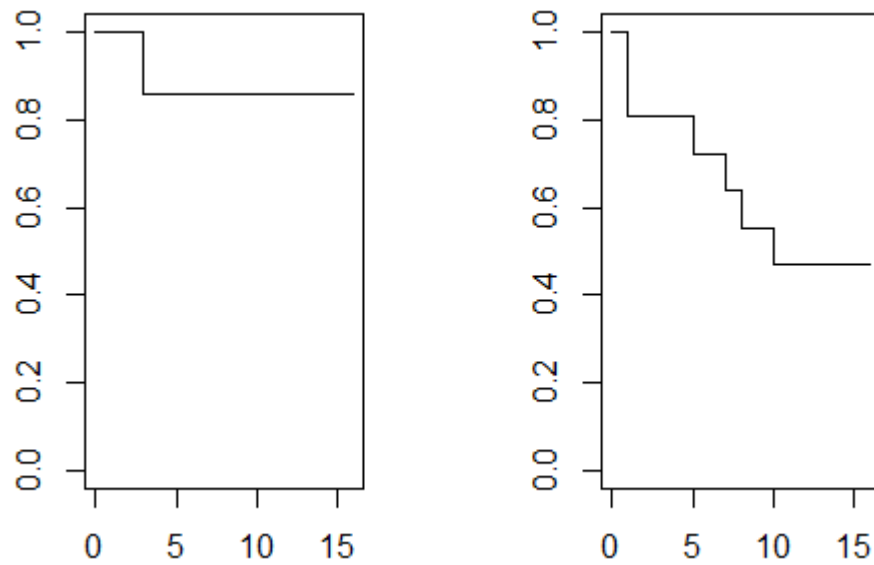
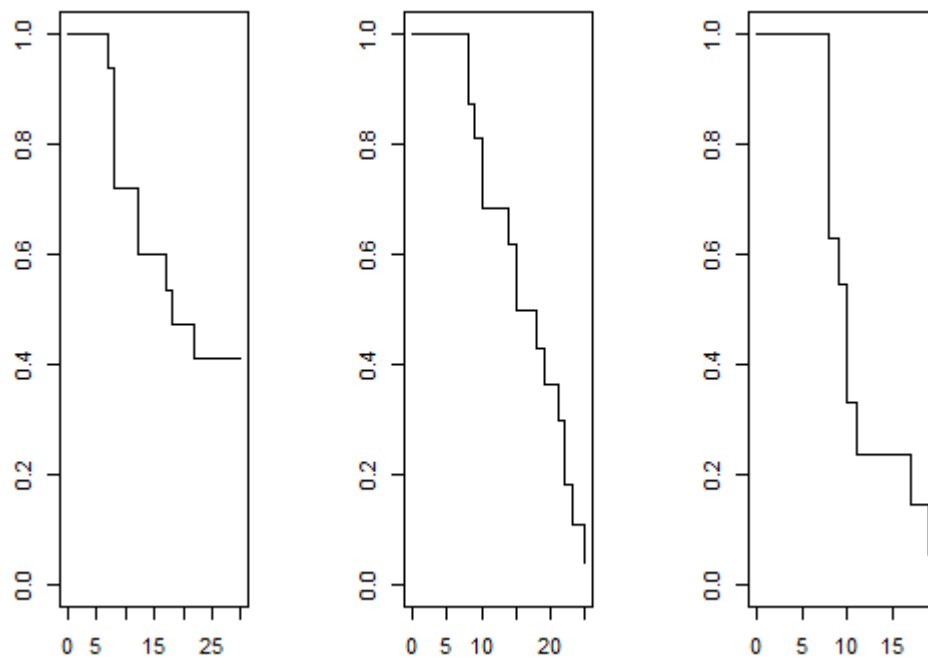


Figura 4 - Curva de Probabilidade $S(t)$ do estimador de Nelson-Aalen para o grupo 2. Da esquerda para a direita temos o grupo 1, grupo 2 e grupo 3.



REFERÊNCIAS

CARVALHO, M. S. et al. Análise de Sobrevivência: Teoria e Aplicações em Saúde, segunda edição, Fiocruz, 2011.