

Introdução à Bioestatística

Marcelo Goulart Correia

Instituto Nacional de Cardiologia

March 7, 2016

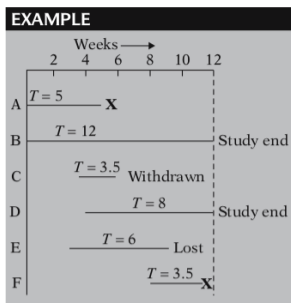
- 1 Análise de sobrevivência
- 2 Censura
- 3 Termos e notações
- 4 Estatísticas descritivas em análise de sobrevivência
- 5 Curva de Kaplan-Meier
- 6 Comparação entre duas ou mais curvas

- Conjunto de análises estatísticas para análise do tempo até a ocorrência de um evento
- Tempo → Anos, dias, meses, ...
- Evento → Morte, Doença, Recaída, ...
- Pode-se analisar mais de um evento (Risco competitivo)
- Variável tempo → Tempo de sobrevivência
- Evento → Falha

- Exemplos

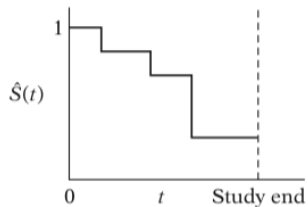
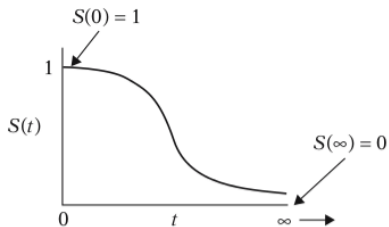
- Tempo (anos) até doença cardíaca em pacientes livres de doença
- Tempo (anos) até morte para pessoas idosas (60+)
- Tempo (meses) até morte para pacientes transplantados
- Tempo (semanas) até remissão em pacientes com leucemia

- Não se tem a informação exata do tempo de sobrevivência de um indivíduo
- Causas:
 - O estudo termina sem que o indivíduo passe por um evento
 - Perda de informação durante o follow-up
 - A pessoa abandona ou **morre** durante o estudo



- $T \rightarrow$ Tempo de sobrevivência de uma pessoa
- $t \rightarrow$ Um valor específico dentro de T
- $\delta \rightarrow$ Variável aleatória (0,1)
 - 0 = Censura
 - 1 = Falha
- $S(t) \rightarrow$ Função de sobrevivência (Probabilidade de uma pessoa sobreviver ao longe de um tempo t)
- $h(t) \rightarrow$ Função de risco (ou taxa de falha condicional)

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (1)$$



- Utilização para função de risco:
 - Entender o comportamento da taxa de falha condicional
 - Entender o formato específico do modelo
 - Modelos matemáticos para análise de sobrevivência são frequentemente escritos a partir da função de risco

- **Objetivos**

- Estimar e interpretar as funções de sobrevivência e de risco
- Comparar as funções de sobrevivência e de risco
- Inferir uma relação das variáveis explicatórias com o tempo de sobrevivência

- Taxa média de risco

$$\bar{h} = \frac{\#Falhas}{\sum_{i=1}^n t_i} \quad (2)$$

- Tempo médio de sobrevivência

$$\bar{T} = \frac{\sum \text{Tempo de sobrevida}}{n} \quad (3)$$

- As estatísticas fornecem uma medida geral e não ao longo do tempo

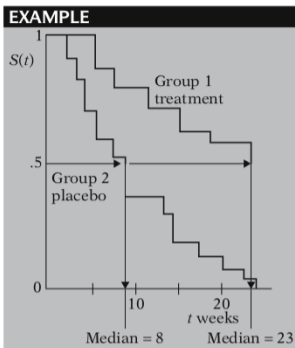
- Exemplo
 - Comparar dois grupos (tratamento e placebo) para verificar o tempo médio (semanas) de sobrevivência e a taxa média de risco para pacientes que vão à remissão em tratamento de leucemia.

Tratamento	Placebo
6	1
6	1
6	2
7	2
10	3
13	4
16	4
22	5
23	5
6+	8
9+	8
10+	8
11+	8
17+	11
19+	11
20+	12
25+	12
32+	15
32+	17
34+	22
35+	23

- $\bar{h}_{Trat} = \frac{9}{6+6+6+7+\dots+22+23} = 0,025$
- $\bar{h}_{Ctrl} = \frac{21}{1+1+2+2+\dots+22+23} = 0,115$
- $\bar{T}_{Trat} = \frac{6+6+6+7+\dots+34+35}{21} = 17,1$ semanas
- $\bar{T}_{Ctrl} = \frac{1+1+2+2+\dots+22+23}{21} = 8,6$ semanas
- Aparentemente tratamento é mais efetivo que o placebo

- Exercício

- Curva de sobrevida estimada



- Mediana do tempo de sobrevivência
- Kaplan-Meier

- Outros fatores importantes:
 - Importante considerar fatores de confundimento e interação
 - Os problemas de modelagem em análise de sobrevivência são análogos aos de modelos de regressão
 - O variável desfecho (tempo até ocorrência de evento) para análise de sobrevivência é diferente dos modelos de regressão
 - Medida de efeito utilizada em análise de sobrevivência → Razão de risco (HR)

- Utilizado para a estimativa e construção gráfica da curva de sobrevivência
- As probabilidades de sobrevivência estimada são calculadas através da fórmula do limite do produto

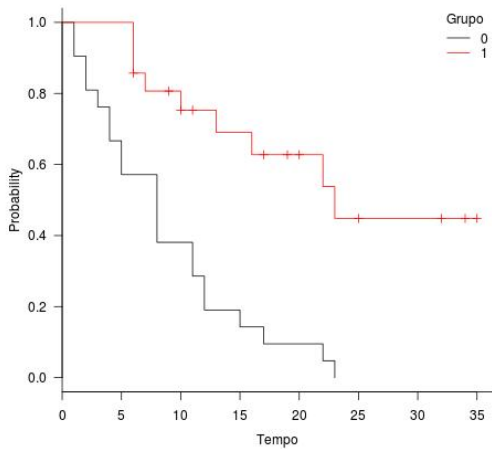
- Objetivo

- Mostrar como são feitos os cálculos da probabilidade de sobrevivência
- Interpretar um gráfico com uma ou mais curvas de sobrevida
- Introduzir o conceito de comparação de mais curvas e suas alternativas

- Exemplo
 - Comparar dois grupos (tratamento e placebo) para verificar o tempo médio (semanas) de sobrevivência e a taxa média de risco para pacientes que vão à remissão em tratamento de leucemia.

Tratamento	Placebo
6	1
6	1
6	2
7	2
10	3
13	4
16	4
22	5
23	5
6+	8
9+	8
10+	8
11+	8
17+	11
19+	11
20+	12
25+	12
32+	15
32+	17
34+	22
35+	23

- $\bar{h}_{Trat} = \frac{9}{6+6+6+7+\dots+22+23} = 0,025$
- $\bar{h}_{Ctrl} = \frac{21}{1+1+2+2+\dots+22+23} = 0,115$
- $\bar{T}_{Trat} = \frac{6+6+6+7+\dots+34+35}{21} = 17,1$ semanas
- $\bar{T}_{Ctrl} = \frac{1+1+2+2+\dots+22+23}{21} = 8,6$ semanas
- $\bar{h}_{Ctrl}/\bar{h}_{Trat} = 4,6$



```
Call: survfit(formula = Surv(Tempo, Remissao == 1) ~ Grupo, data = Dados_Exemplo,
na.action = na.omit, conf.type = "log-log")
```

Grupo=0

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	21	2	0.9048	0.0641	0.67005	0.975
2	19	2	0.8095	0.0857	0.56891	0.924
3	17	1	0.7619	0.0929	0.51939	0.893
4	16	2	0.6667	0.1029	0.42535	0.825
5	14	2	0.5714	0.1080	0.33798	0.749
8	12	4	0.3810	0.1060	0.18307	0.578
11	8	2	0.2857	0.0986	0.11656	0.482
12	6	2	0.1905	0.0857	0.05948	0.377
15	4	1	0.1429	0.0764	0.03566	0.321
17	3	1	0.0952	0.0641	0.01626	0.261
22	2	1	0.0476	0.0465	0.00332	0.197
23	1	1	0.0000	NaN	NA	NA

Grupo=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	21	3	0.857	0.0764	0.620	0.952
7	17	1	0.807	0.0869	0.563	0.923
10	15	1	0.753	0.0963	0.503	0.889
13	12	1	0.690	0.1068	0.432	0.849
16	11	1	0.627	0.1141	0.368	0.805
22	7	1	0.538	0.1282	0.268	0.747
23	6	1	0.448	0.1346	0.188	0.680

- time → Instante de tempo
- n.risk → Pacientes em risco
- n.event → Pacientes em remissão
- survival → Probabilidade de sobrevivência do instante de tempo
- std.err → Erro padrão da probabilidade
- lower (upper) 95% CI → Intervalo de confiança da probabilidade

- Exercício

- Existem alguns testes estatísticos que auxiliam na verificação da diferença entre as curvas de sobrevida:
 - **Log-Rank**
 - Wilcoxon
 - Tarone-Ware
 - Peto
 - Fleimington-Harrigton

- Teste Log-Rank (Mantel-Haenszel)
 - Utilizada para a comparação entre as curvas de sobrevivência analisadas
 - H_0 : Não existem diferenças entre as curvas de sobrevivência analisadas
 - H_A : Existem diferenças entre as curvas de sobrevivência analisadas

- Metodologia:
 - Calcular os valores esperados

$$e_{1j} = \frac{n_{1j}}{n_{1j} + n_{2j}} * (m_{1j} + m_{2j}) \quad (4)$$

$$e_{2j} = \frac{n_{2j}}{n_{1j} + n_{2j}} * (m_{1j} + m_{2j}) \quad (5)$$

- Verificar os valores observados
- Calcular o Log-Rank

$$\text{Log-Rank} = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)} \quad (6)$$

- Aproximando o Log-Rank para Qui-Quadrado:

$$\chi^2 \approx \sum_i^{\#grupos} \frac{(O_i + E_i)^2}{E_i} \quad (7)$$

- Exemplo
 - Comparar dois grupos (tratamento e placebo) para verificar o tempo médio (semanas) de sobrevivência e a taxa média de risco para pacientes que vão à remissão em tratamento de leucemia.

Tratamento	Placebo
6	1
6	1
6	2
7	2
10	3
13	4
16	4
22	5
23	5
6+	8
9+	8
10+	8
11+	8
17+	11
19+	11
20+	12
25+	12
32+	15
32+	17
34+	22
35+	23

- $\text{Log-Rank} = \frac{(10,26)^2}{6,2685} = 16,793$
- $\chi^2 \approx \frac{(-10,26)^2}{19,26} + \frac{(10,26)^2}{10,74} = 15,276$
- Valor crítico = 5,024
- Valor de p = 0,0000417 (GL = 1)
- Existem diferenças entre as curvas analisadas

- Exercício