ANÁLISE DE SOBREVIVÊNCIA

Airlane P. Alencar – IME-USP

Alessandra C. Goulart – FM-USP

Gisela Tunes Silva – IME-USP

Objetivo

- Estudar o tempo desde um instante inicial até a ocorrência de um evento (falha).
- Estudar o tempo de sobrevida de um paciente a partir de um instante inicial, por exemplo após o primeiro AVC.
- O que queremos saber?

  O tempo médio de vida após um AVC para homens, mulheres, dependendo do tipo de AVC.

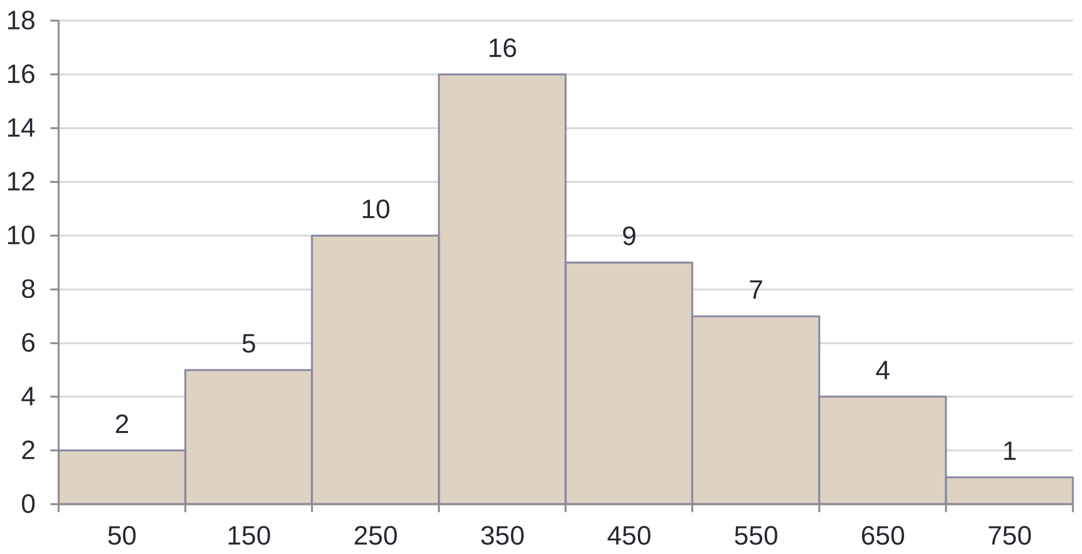
  Qual a prob de sobreviver 1 ano, 2 anos pós AVC?

- Referências
  - Colosimo e Giolo
  - Kleinbaum e Klein

Estimar a prob. de sobrevida

- Tempo de sobrevida sem censura

Tempo	Frequência
0-100	2
100-200	5
200-300	10
300-400	16
400-500	9
500-600	7
600-700	4
700-800	1
800-900	0
Total	54



Fonte: Colosimo e Giolo p.33

Probabilidade de Sobrevida

- P(T> 100) = prob. de viver mais que 100 dias

  não morreu antes de 100

Tempo	Frequência	Suscetíveis no início
0-|100	2	54
100-|200	5	52
200-|300	10	47
300-|400	16	37
400-|500	9	21
500-|600	7	12
600-|700	4	5
700-|800	1	1
800-|900	0	0
Total	54	

Tempo	P(S>t)	
0	1	
100	0.963	= 52/54
200	0.870	= 47/54
300	0.685	
400	0.389	
500	0.222	
600	0.093	
700	0.019	
800	0.000	

Taxa de falha

- Sem censura, a taxa de falha (λ) em um intervalo é quantos falharam com relação a quantos estavam suscetíveis com relação à duração do intervalo

Tempo	Frequência	Suscetíveis no início	Taxa de falha no intervalo
0-|100	2	54	0.037
100-|200	5	52	0.096
200-|300	10	47	0.213
300-|400	16	37	0.432
400-|500	9	21	0.429
500-|600	7	12	0.583
600-|700	4	5	0.800
700-|800	1	1	1.000
800-|900	0	0	
Total	54		

Taxa de falha(400-500)=

9/21=42,9%

Se já sobreviveu até o tempo 400, a chance de falhar nesse intervalo de 100 horas é 42,9%, ie, taxa=42,9%/100 h

Censura

- Poderia propor modelos como os de regressão e análise de variância para a variável resposta Tempo de vida

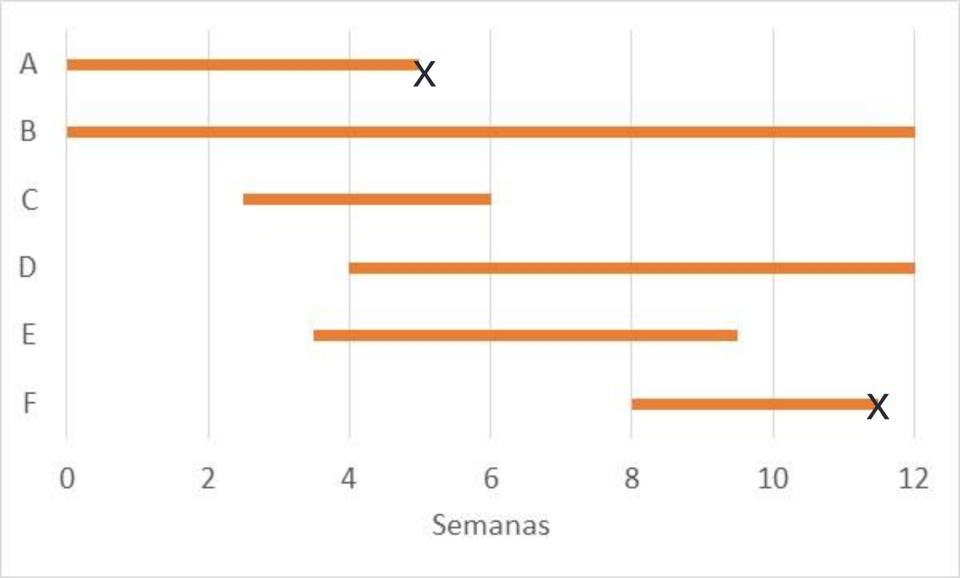
- Mas se observei que um paciente viveu mais que 800 dias e não sei quando morreu, tenho que inclui-lo na análise! Como?

Censura?

- O estudo terminou e não observou-se a falha.
- Perda de seguimento (follow up)
- Sei que estava vivo até certo tempo (T>x)
- A pessoa sai do estudo por ocorrência de outro evento. Ex: efeito colateral, transplante, óbito quando não for o evento de interesse

Tempos iniciais e de sobrevida

Pessoa	Tempo	Falha
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1



- Tempos: 5, 12+, 3.5+, 8+, 6+, 3.5

Tipos de censura

- Censura à direita é a mais usual:

  Tempo de sobrevida (T) >= tempo observado O.

- Censura à esquerda: T<= O

  Follow up até pessoas serem HIV+. Fez teste em t e deu positivo então sei que T<t.

- Censura intervalar: Só sei que t1<=T<=t2

  Fiz testes nos instantes t1 e t2

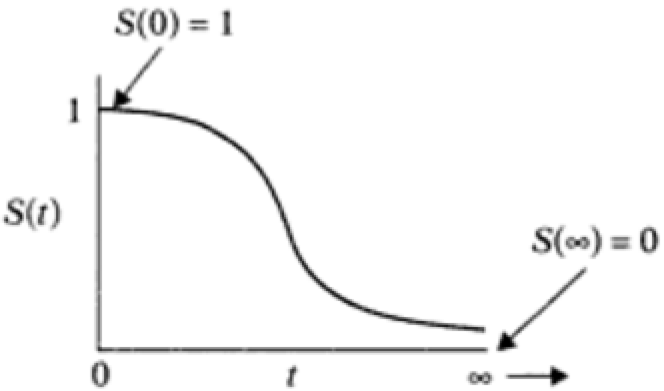
Tipos de censura

- Aleatória: Tempo de falha (T) e de censura (C) aleatórios e observamos t=min(T,C).
- Censura tipo I: Estudo acaba após certo tempo e temos r falhas e no final do estudo temos n-r censuras.
- Censura tipo II: r é fixado e só os menores r tempos são observados e todos os outros tempos são censurados. O maior tempo observado é $t_{(r)}$.
- Independente: e se a pessoa com melhor prognóstico sempre larga o estudo?

- Colosimo e Giolo: exercícios p.26

Curva de sobrevivência

- Apresenta a probabilidade de sobrevivência = função de sobrevivência = S(t)= P(T > t)



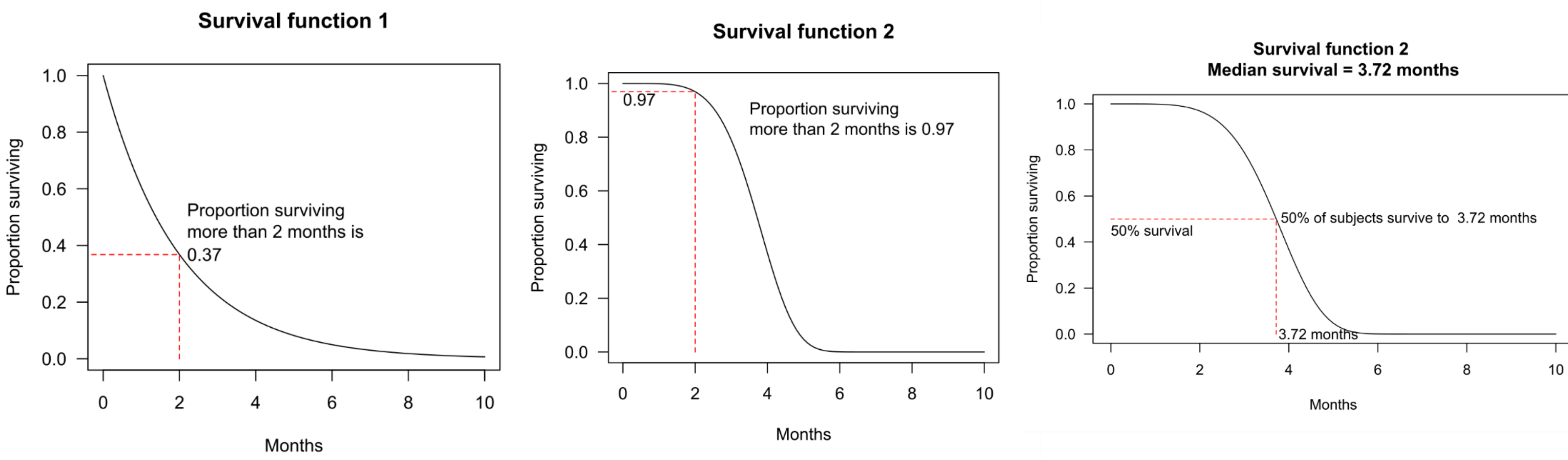
Função não crescente

com S(0)=1 e que tende

a 0 (sem cura= falha ocorre)

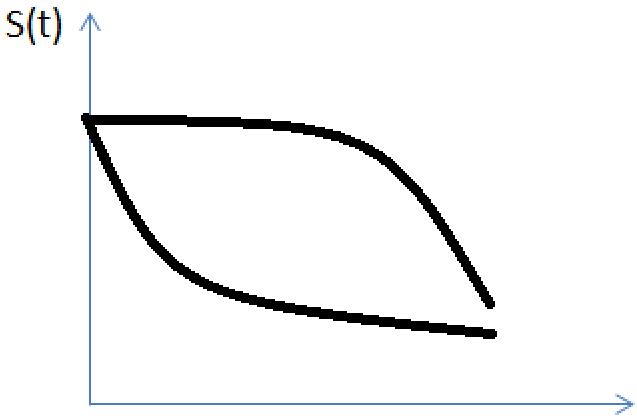
- S(1) = P(T>1) >= S(2)...
- Vamos estimar essa curva usando os dados

Curva de sobrevivência



Curva de sobrevivência

- Qual curva de sobrevida você prefere para procedimento após sua cirurgia?



Taxa de falha h(t)

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

- h(t)>=0 e não tem limite máximo
- h(t) é uma taxa que mede o potencial instantâneo
- Ex: Constante para saudáveis (aumenta com idade)

  Maior logo após um AVC e depois decresce

  Aumenta para pacientes com leucemia

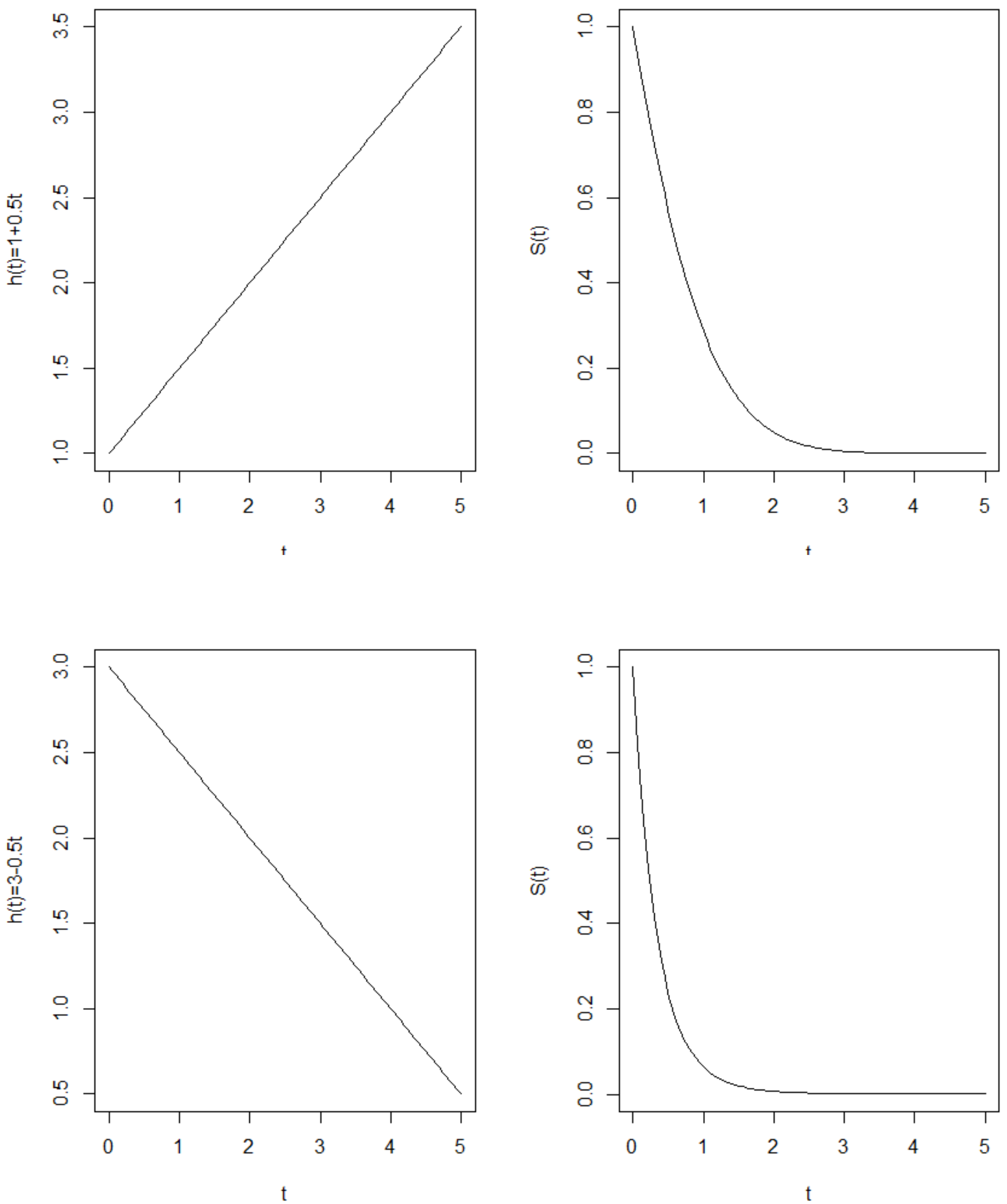
Taxa e Curva de Sobrevida

- S(t) e h(t) são tais que

$$S(t) = \exp\left[-\int_0^t h(u)du \right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)} \right]$$

- h(t) mede o quanto varia S(t)



$$S(t) = \exp\left[-\int_0^t h(u)\,du\right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right]$$

Estimativa de S(t) – Kaplan Meier

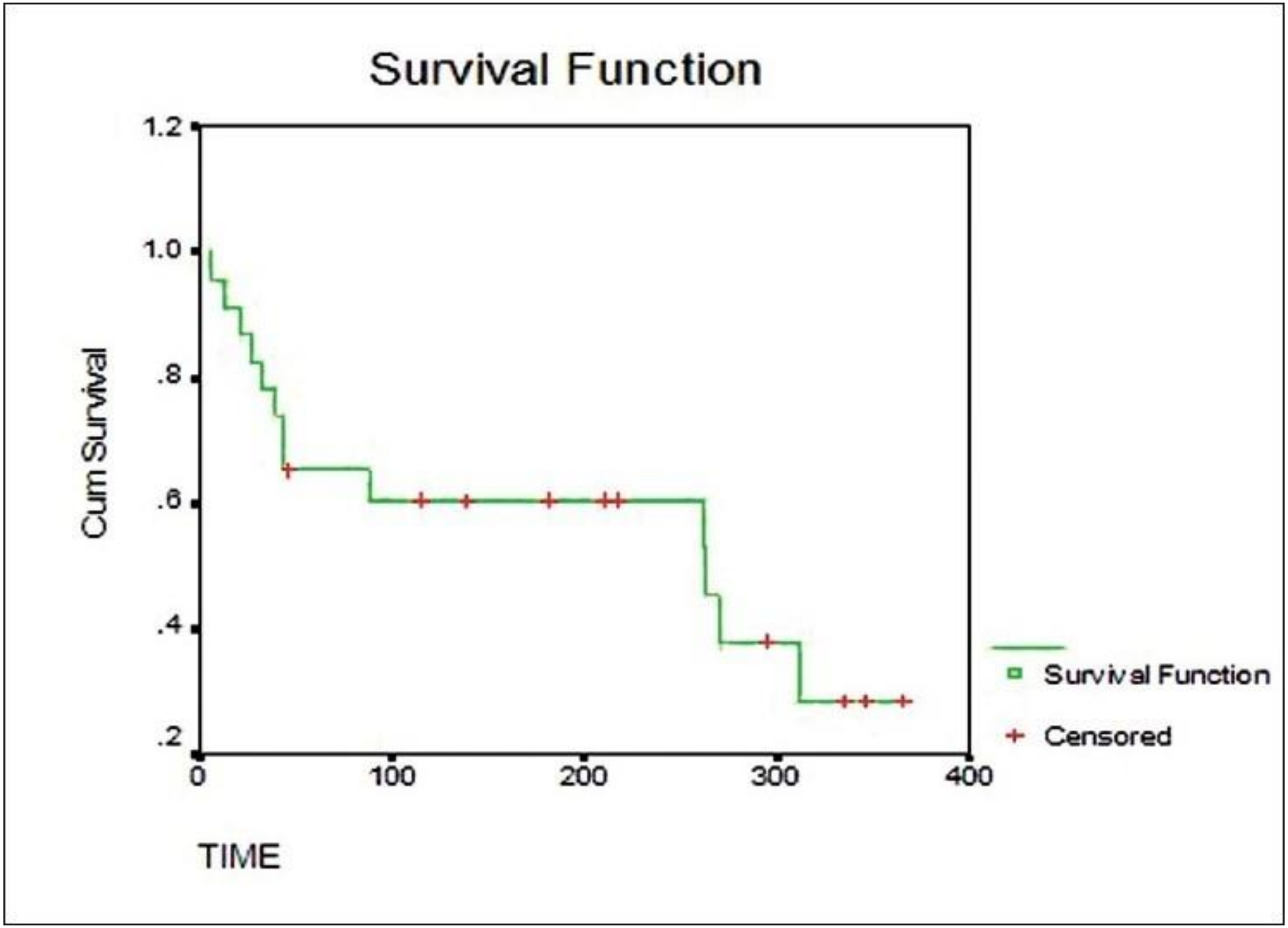
- **Goel et al. 2010. Understanding survival analysis: Kaplan-Meier estimate.** Int J Ayurveda Res. 2010 Oct-Dec; 1(4): 274–278.
-
- 6, 12, 21, 27, 32, 39, 43, 43, 46+, 89, 115+, 139+, 181+, 211+, 217+, 261, 263, 270, 295+, 311, 335+, 346+, 365+

- t(i) é o i-ésimo tempo de falha ordenado.

Estimador Kaplan Meier (1958)

Time of event (t)	No. of Pt. died (d)	Live at the start of the day (n)	Estimated probability		Probability of survivors at the end of time (L)
			death (d/n)	survival (1 - d/n)	
6	1	23	0.0435	0.9565	0.9565
12	1	22	0.0455	0.9545	0.9565 × 0.9545 = 0.9130
21	1	21	0.0476	0.9524	0.9130 × 0.9523 = 0.8695
27	1	20	0.0500	0.9500	0.8695 × 0.9500 = 0.8260
32	1	19	0.0526	0.9474	0.7826
39	1	18	0.0556	0.9444	0.7391
43	2	17	0.1176	0.8824	0.6522
89	1	14	0.0714	0.9286	0.6056
261	1	8	0.125	0.875	0.5299
263	1	7	0.1429	0.8571	0.4542
270	1	6	0.1667	0.8333	0.3785
311	1	4	0.25	0.75	0.2839

The time 't' for which the value of 'L' i.e. total probability of survival at the end of a particular time is 0.50 is called as median survival time. The estimates obtained are invariably expressed in graphical form. The graph plotted between estimated survival probabilities/estimated survival percentages (on Y axis) and time past after entry into the study (on X axis) consists of horizontal and vertical lines.[4] The survival curve is drawn as a step function: the proportion surviving remains unchanged between the events, even if there are some intermediate censored observations. It is incorrect to join the calculated points by sloping lines (Figure 1).


Survival Function
Cum Survival
1.2
1.0
.8
.6
.4
.2
0
100
200
300
400
TIME
Survival Function
Censored

Objetivo

- Estimar curvas de sobrevida levando-se em conta a censura (KK)

Grupo 1 - Tratamento - n=21 - 9 falhas e 12 censuras

6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+

Grupo 2 - Placebo - n=21 - 21 falhas

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Dados formato longo

ind	t	Falha	Grupo
1	6	1	1
2	6	1	1
3	6	1	1
4	7	1	1
5	10	1	1
6	13	1	1
7	16	1	1
8	22	1	1
9	23	1	1
10	6	0	1
11	9	0	1
12	10	0	1
13	11	0	1
14	17	0	1
15	19	0	1
16	20	0	1
17	25	0	1
18	32	0	1
19	32	0	1
20	34	0	1
21	35	0	1
1	1	1	2
2	1	1	2
3	2	1	2
4	2	1	2
5	3	1	2
6	4	1	2
7	4	1	2

Estimativa de S(t) – Kaplan Meier

- 6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+

			censura	expostos	
i	t(i)	falhas	[t(i), t(i+1))	sobreviveram até t(i) (# T>=t(i))	
	0	0	0	21	sobrevivem >=0 semanas
1	6	3	1	21	sobrevivem >=6 semanas
2	7	1	1	17	sobrevivem >=7 semanas
3	10	1	2	15	sobrevivem >=10 semanas
4	13	1	0	12	sobrevivem >=13 semanas
5	16	1	3	11	sobrevivem >=16 semanas
6	22	1	0	7	sobrevivem >=22 semanas
7	23	1	5	6	sobrevivem >=23semanas
	Total	9	12		

Probabilidade de falha e sobrevivência

- Em [t(i), t(i+1))
- P(morrer em [6,7))= 3/21
- P(sobrevivem ao [6,7))= 1-3/21

			censura	expostos	Em [t(i), t(i+1))	
i	t(i)	falhas	[t(i), t(i+1))	(# T>=t(i))	P(morrer)	P(sobr)=1-P(m)
	0	0	0	21	0	1
1	6	3	1	21	0.143	0.857
2	7	1	1	17	0.059	0.941
3	10	1	2	15	0.067	0.933
4	13	1	0	12	0.083	0.917
5	16	1	3	11	0.091	0.909
6	22	1	0	7	0.143	0.857
7	23	1	5	6	0.167	0.833
	Total	9	12			

Estimar S(t)=P(T>t) - KM

- P(T>6)= 0.857
- P(T>7)= P(T>7|T>6)P(T>6)= 0.857*0.941

			censura	expostos	Em [t(i), t(i+1))		
i	t(i)	falhas	[t(i), t(i+1))	(# T>=t(i))	P(morrer)	P(sobr)=1-P(m)	S^(t)= S estimada
	0	0	0	21	0	1	1
1	6	3	1	21	0.143	0.857	0.857
2	7	1	1	17	0.059	0.941	
3	10	1	2	15	0.067	0.933	
4	13	1	0	12	0.083	0.917	
5	16	1	3	11	0.091	0.909	
6	22	1	0	7	0.143	0.857	
7	23	1	5	6	0.167	0.833	
	Total	9	12				

Estimar S(t)=P(T>t) - KM

- P(T>7)= P(T>7|T>6)P(T>6)= 0.857*0.941=0.807
- P(T>10)= P(T>10|T>7)P(T>7)= 0.807*0.933=

			censura	expostos	Em [t(i), t(i+1))		
i	t(i)	falhas	[t(i), t(i+1))	(# T>=t(i))	P(morrer)	P(sobr)=1-P(m)	S^(t)= S estimada
	0	0	0	21	0	1	1
1	6	3	1	21	0.143	0.857	0.857
2	7	1	1	17	0.059	0.941	0.807
3	10	1	2	15	0.067	0.933	
4	13	1	0	12	0.083	0.917	
5	16	1	3	11	0.091	0.909	
6	22	1	0	7	0.143	0.857	
7	23	1	5	6	0.167	0.833	
	Total	9	12				

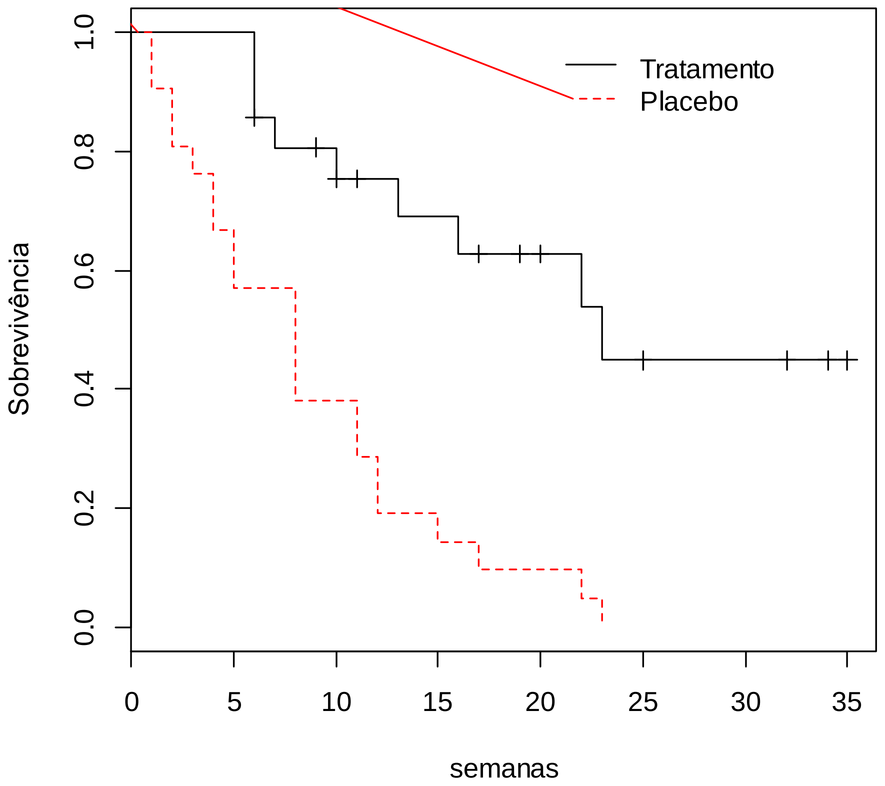
Estimar S(t) - KM

			censura	expostos	Em [t(i), t(i+1))		
i	t(i)	falhas	[t(i), t(i+1))	(# T>=t(i))	P(morrer)	P(sobr)=1-P(m)	S^(t)= S estimada
	0	0	0	21	0	1	1
1	6	3	1	21	0.143	0.857	0.857
2	7	1	1	17	0.059	0.941	0.807
3	10	1	2	15	0.067	0.933	0.753
4	13	1	0	12	0.083	0.917	0.690
5	16	1	3	11	0.091	0.909	0.627
6	22	1	0	7	0.143	0.857	0.538
7	23	1	5	6	0.167	0.833	0.448
	Total	9	12				

- P(T>6)= 0.857
- P(T>7)= P(T>7|T>6)P(T>6)= 0.857*0.941

$$\hat{S}\left(t_{(f)}\right) = \prod_{i=1}^{f} P\left(T > t_{(i)} \mid T > t_{(i-1)}\right)$$

Gráfico



Tratamento
Placebo
Sobrevivência
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
semanas

Propriedades do estimador

O estimador de Kaplan-Meier é

- Não viciado para amostras grandes
- É consistente
- Converge assintoticamente para processo gaussiano
- É o estimador de máxima verossimilhança de S(t)
- Variância do estimador e IC assint. p.41 Fórmula de Greenwood
- Para altos valores de t, o IC pode apresentar valor <0 e pode ser usada transf. de $\hat{S}$(t). Default do R log($\hat{S}$(t)).
- Outro estimador é o de Nelson-Aalen (vide p. 43 Colosimo e Giolo)

Com censura

- Exemplo: Colosimo e Giolo
- Terapia esteróide no tratamento de hepatite viral aguda (Gregory al., 1976)
- 29 pacientes aleatorizados: 14 com esteroide
- Acompanhamento de 29 semanas

Exercício Colosimo e Giolo p39

- Tempos de sobrevida (até a morte ou censura)
- 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+
- Usar comandos no R

Comparação de curvas - Kleinbaum e Klein p.61

- Grupo 1 - Tratamento - n=21 - 9 falhas e 12 censuras
- 6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+
- Grupo 2 - Placebo - n=21 - 21 falhas
- 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

- Tabela com os tempos de falha e número de óbitos (m) e suscetíveis (n) segundo o grupo

Comparação de curvas - Kleinbaum e Klein p.61

- Qual o valor esperado para cada tempo e grupo se a sobrevida fosse igual nos 2 grupos?

- Esperado no grupo 1:

$$e1f = \left(\frac{m1f + m2f}{n1f + n2f} \right) \cdot n1f$$

Mortalidade igual nos 2 grupos em t(f)

t(f)	m1f	m2f	n1f	n2f
1	0	2	21	21
2	0	2	21	19
3	0	1	21	17
4	0	2	21	16
5	0	2	21	14
6	3	0	21	12
7	1	0	17	12
8	0	4	16	12
10	1	0	15	8
11	0	2	13	8
12	0	2	12	6
13	1	0	12	4
15	0	1	11	4
16	1	0	11	3
17	0	1	10	3
22	1	1	7	2
23	1	1	6	1
soma	9	21		

Kleinbaum e Klein

Estatística Log Rank

t(f)	m1f	m2f	n1f	n2f	e1f	e2f	m1f-e1f	m2f-e2f
1	0	2	21	21	(2/42) 21=1	(2/42) 21=1	-1.00	1.00
2	0	2	21	19	(2/40) 21=1.1	(2/40) 19=1	-1.05	1.05
3	0	1	21	17	(1/38) 21=0.6	(1/38) 17=0.4	-0.55	0.55
4	0	2	21	16	(2/37) 21=1.1	(2/37) 16=0.9	-1.14	1.14
5	0	2	21	14	(2/35) 21=1.2	(2/35) 14=0.8	-1.20	1.20
6	3	0	21	12	(3/33) 21=1.9	(3/33) 12=1.1	1.09	-1.09
7	1	0	17	12	(1/29) 17=0.6	(1/29) 12=0.4	0.41	-0.41
8	0	4	16	12	(4/28) 16=2.3	(4/28) 12=1.7	-2.29	2.29
10	1	0	15	8	(1/23) 15=0.7	(1/23) 8=0.3	0.35	-0.35
11	0	2	13	8	(2/21) 13=1.2	(2/21) 8=0.8	-1.24	1.24
12	0	2	12	6	(2/18) 12=1.3	(2/18) 6=0.7	-1.33	1.33
13	1	0	12	4	(1/16) 12=0.8	(1/16) 4=0.3	0.25	-0.25
15	0	1	11	4	(1/15) 11=0.7	(1/15) 4=0.3	-0.73	0.73
16	1	0	11	3	(1/14) 11=0.8	(1/14) 3=0.2	0.21	-0.21
17	0	1	10	3	(1/13) 10=0.8	(1/13) 3=0.2	-0.77	0.77
22	1	1	7	2	(2/9) 7=1.6	(2/9) 2=0.4	-0.56	0.56
23	1	1	6	1	(2/7) 6=1.7	(2/7) 1=0.3	-0.71	0.71
soma	9	21			19.25	10.75	-10.25	10.25

Teste Log rank

- A estatística de teste compara os valores observados e os valores esperados em um grupo

$$estat = \frac{(O_2 - E_2)^2}{Var(O_2 - E_2)} = \frac{\left(\sum_f (m_{2f} - e_{2f})\right)^2}{Var(O_2 - E_2)}$$

$$Var(O_i - E_i) = \sum_f \frac{n_{1f} n_{2f} (m_{1f} + m_{2f})(n_{1f} + n_{2f} - m_{1f} - m_{2f})}{(n_{1f} + n_{2f})^2 (n_{1f} + n_{2f} - 1)}$$

i= 1 ou 2

- H0: as curvas são iguais
- Sob H0, $estat \sim \chi_1^2$

Teste log rank - exemplo

- Estat= (10.25)^2/6.26 = 16.7929
- Valor p = P(quiquad1>16.7929) < 0.0001
- No excel: Valor p=1-DIST.QUIQUA(B77;1;1)
- Conclusão: Há diferença significativa entre as curvas

Log rank para g grupos

- A estatística envolve a diferença entre valores observados e esperados para todos os grupos mas precisa calcular a matriz de variâncias e covariâncias dessas diferenças.
- Sob H0
- Esse teste é assintótico, ou seja, essa distribuição vale para n grande

$$estat \sim \chi^2_{g-1}$$

Outros testes

- Outros testes semelhantes ao teste log rank incluem pesos wf para cada tempo t(f).
- Estatística ponderada

$$= \frac{\left(\sum_f w_f \left(m_{if} - e_{if}\right)\right)^2}{\operatorname{var}\left(\sum_f w_f \left(m_{if} - e_{if}\right)\right)}$$

- Teste de Wilcoxon: wf=nf
  - Maior peso para os tempos iniciais

Outros testes

Teste	Peso
Log rank	1
Wilcoxon	n_f
Tarone-Ware	$\sqrt{n_f}$
Peto	Sobrevivência estimada combinada nos 2 grupos
Flemington- Harrington	$\hat{S}\left(t_{(f-1)}\right)^p \left[1-\hat{S}\left(t_{(f-1)}\right)\right]^q$

Teste estratificado

- Compara 2 curvas de sobrevida controlando por uma variável categorizada (G estratos)
- Calcula as diferenças Oi-Ei para cada estrato e soma essas diferenças.
- Elas devem estar no mesmo sentido nos vários estratos.
- Sob H0

$$estat \sim \chi^2_{G-1}$$

Intervalo de confiança para S(t)

- O estimador de Kaplan Meier para S(t) tem distribuição normal assintótica.
- Intervalo de confiança para S(t)

- Fórmula de Greenwood:

$$IC = \left[\hat{S}(t) \mp z\sqrt{v\hat{a}r\left(\hat{S}(t)\right)} \right]$$

$$v\hat{a}r\left(\hat{S}(t)\right) = \left(\hat{S}(t)\right)^2 \sum_{f:t_{(f)}\leq t} \left[\frac{m_f}{n_f\left(n_f - m_f\right)} \right]$$

- Obs: Para valores extremos de t, este IC pode apresentar limites negativos ou maiores que 1. Kalbfleish e Prentice (1980) sugerem usar U^(u)=log[-log(^S(t))] e sua variância assintótica para construir IC. (Colosimo e Giolo, p.42).

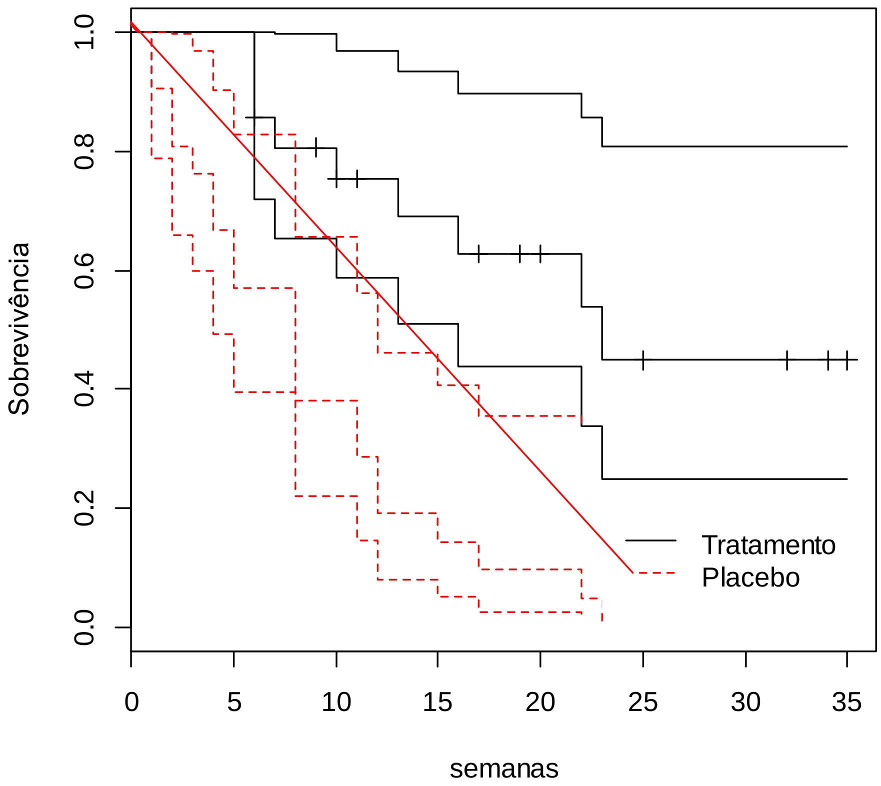
Exemplo KK

- IC – Grupo Tratamento

Grupo=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	21	3	0.857	0.0764	0.720	1.000
7	17	1	0.807	0.0869	0.653	0.996
10	15	1	0.753	0.0963	0.586	0.968
13	12	1	0.690	0.1068	0.510	0.935
16	11	1	0.627	0.1141	0.439	0.896
22	7	1	0.538	0.1282	0.337	0.858
23	6	1	0.448	0.1346	0.249	0.807

Kaplan Meier com ICs



No R

```
d=read.csv("C:/Users/Lane/Dropbox/2018/surv/data/kk1.csv", sep=";")
names(d)
# Grupo: 1=tratamento e 2=Placebo
d$Grupo <- factor(d$Grupo, labels = c("tratamento", "placebo"))

library(survival)
fit <- survfit(Surv(t, Falha) ~ Grupo, data=d)
summary(fit)
plot(fit, lty = 1:2, col=1:2, xlab= "semanas", ylab="Sobrevivência")
legend(20, 1, c("Tratamento", "Placebo"), lty = 1:2, col=1:2, box.col="white")

plot(fit, conf.int=T, lty = 1:2, col=1:2, xlab= "semanas", ylab="Sobrevivência")
legend(23, .2, c("Tratamento", "Placebo"), lty = 1:2, col=1:2, box.col="white")

survdiff(formula = Surv(t, Falha) ~ Grupo, data=d)
```

- survdiff(formula = Surv(t, Falha) ~ Grupo, data = d)

- N Observed Expected (O-E)^2/E (O-E)^2/V
- Grupo=tratamento 21 9 19.3 5.46 16.8
- Grupo=placebo 21 21 10.7 9.77 16.8

Bibliografia

- Kleinbaum e Klein. Survival analysis – a self-learning text.

- Colosimo e Giolo. Análise de sobrevivência aplicada.