Análise de Sobrevivência Aplicada à Saúde

Lupércio França Bessegato
Dep. Estatística/UFJF

Roteiro

Estimação Não Paramétrica

Estimadores Não Paramétricos

- Abordagem não paramétrica:
  - √ Não adota qualquer suposição sobre a distribuição de probabilidades do tempo de sobrevivência (T)
- Estimadores:
  - √ Kaplan-Meier:
    - Estima a função de sobrevivência – S(t)
  - √ Nelson-Aalen:
    - Estima a função de risco acumulado – Λ(t)
  - √ Ambos utilizam dados censurados

Estimador de Kaplan-Meier

- Desdobra a condição de sobreviver até o tempo *t*
  - √ Sequência de elementos independentes, caracterizando a sobrevivência em cada intervalo de tempo anterior a *t*
  - √ A probabilidade de sobreviver ao tempo t é o produto das probabilidades de chegar até cada um dos tempos anteriores

    S(5) = P{T>5} = P{T>1; T>5} = P{T>1} P{T>5 | T>1}

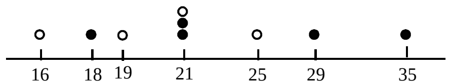
    - Sobrevivência por 5 semanas: necessário sobreviver à 1ª semana e depois, à 5ª semana, sabendo-se que sobreviveu à 1ª
  - √ Estimador produto (ou estimador limite produto)

Exemplo

- Tempo: 25+
  - √ Até a data de sua saída (t < 25)
    - Numerador (sobrevivente) e denominador (grupo sob risco)
  - √ Após sua saída (t > 25)
    - Nem numerador (sem acompanhamento) nem denominador (não pertence mais a grupo sob risco)

Exemplo

- Tempos:
  - √ 16+, 18, 19+, 21, 21+, 21, 25+, 29, 35
  - √ 4 desfechos e 4 censuras



Tempo	Sob risco	Desfechos	Censuras	Sobreviventes	S(t)	Cálculo
0	9	0	0	9	1,000	9/9
16	9	0	1	9	1,000	9/9
18	8	1	0	7	0,875	(7/8)
19	7	0	1	6	0,875	(7/8)(7/7)
21	6	2	1	4	0,583	(4/6)(7/8)
25	3		1	3	0,583	(4/6)(7/8)(3/3)
29	2	1	0	1	0,292	(1/2)(4/6) (7/8)
35	1	1	0	0	0	(0/1) (1/2)(4/6) (7/8)

Estimador Produto

- Amostra com *n* indivíduos e *m* diferentes tempos de ocorrência dos eventos:
  - √ $t_1 < t_2 < \ldots < t_m$
  - √ $R(t_j)$: total de pessoas sob risco no tempo tj
  - √ $\Delta N(t_j)$: número de eventos ocorridos precisamente em t_j
  - √ $R(t_j) - \Delta N(t_j)$: número de sobreviventes no tempo t_j

$$\mathrm{P}\left\{\text{sobrevivência ao tempo } t_j\right\} = \frac{R(t_j) - \Delta N(t_j)}{R(t_j)}$$

√ Como os eventos são independentes S(t) é o produto da probabilidades de sobrevivência a cada tempo $t_j \leq t$

$$\hat{S}_{KM}(t) = \left(\frac{R(t_1) - \Delta N(t_1)}{R(t_1)}\right) \times \left(\frac{R(t_2) - \Delta N(t_2)}{R(t_2)}\right) \times \ldots \left(\frac{R(t_m) - \Delta N(t_m)}{R(t_m)}\right)$$

$$\hat{S}_{KM}(t) = \prod_{t_j \leq t} \left(\frac{R(t_j) - \Delta N(t_j)}{R(t_j)}\right)$$

Exemplo

- Sobrevivência ao diagnóstico de Aids
  - √ Pacientes com tempo de observação menor que 90 dias
    - Óbitos: 15
    - Censuras: 6
  - √ Banco de dados: ipec.csv

```
> ipec90 <- ipec[ipec$tempo < 90,]
> Surv(ipec90$tempo, ipec90$status)
 [1] 60  84  25+ 54  80+ 37  18  29  50+ 83  80  81+ 35  52  21  40  22  85+ 39
[20] 16  21+
```


t_j	$R(t)$	$\Delta N(t)$	$R(t) - \Delta N(t)$	$\hat{S}_{KM}(t)$	$\prod_{t_j \leq t}\left(\frac{R(t_j)-\Delta N(t_j)}{R(t_j)}\right)$
16	21	1	20	0,9524	20/21
18	20	1	19	0,9048	(0,9524) (19/20)
21	19	1	18	0,8571	(0,9048) (18/19)
22	17	1	16	0,8067	(0,8571) (16/17)
29	15	1	14	0,7529	(0,8067) (14/15)
35	14	1	13	0,6992	(0,7529) (13/14)
37	13	1	12	0,6454	(0,6992) (12/13)
39	12	1	11	0,5916	(0,6454) (11/12)
40	11	1	10	0,5378	(0,5916) (10/11)
52	9	1	8	0,4781	(0,5378) (8/9)
54	8	1	7	0,4183	(0,4781) (7/8)
60	7	1	6	0,3585	(0,4183) (6/7)
80	6	1	5	0,2988	(0,3585) (5/6)
83	3	1	2	0,1992	(0,2988) (2/3)
84	2	1	1	0,0996	(0,1992) (1/2)

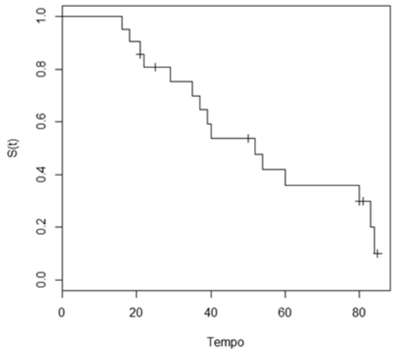
√ Função de sobrevivência estimada pelo método Kaplan-Meier

```
> sobrev.km<-survfit( formula = Surv(tempo, status)~1, data = ipec90)
> summary(sobrev.km)
```

```
time n.risk n.event survival s
  16     21       1   0.9524
  18     20       1   0.9048
  21     19       1   0.8571
  22     17       1   0.8067
  29     15       1   0.7529
  35     14       1   0.6992
  37     13       1   0.6454
  39     12       1   0.5916
  40     11       1   0.5378
  52      9       1   0.4781
  54      8       1   0.4183
  60      7       1   0.3585
  80      6       1   0.2988
  83      3       1   0.1992
  84      2       1   0.0996
```

$$\hat{S}_{KM}(18) = 0,9524 \times \left(\frac{20-1}{20}\right) = 0,9048$$

- Função de sobrevivência dos pacientes com Aids



Função escada:
Salta em cada tempo de ocorrência de evento

√ Símbolos + localizam as censuras

Estimador de Kaplan-Meier para Λ(t)

- Função de risco acumulado:

$$\hat{\Lambda}_{KM}(t) = -\ln \hat{S}_{KM}(t)$$

√ Pode-se estimar qualquer das funções

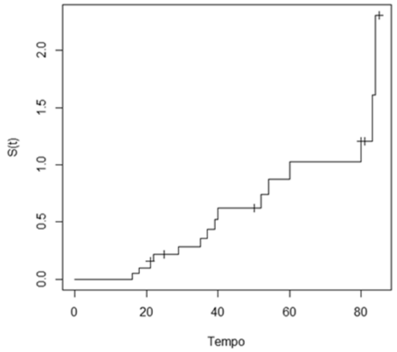
t_j	$R(t)$	$\Delta N(t)$	$R(t) - \Delta N(t)$	$\hat{S}_{KM}(t)$	$\hat{\Lambda}_{KM}$
16	21	1	20	0,9524	–ln(0,9524) = 0,0488
18	20	1	19	0,9048	–ln(0,9048) =0,1001
21	19	1	18	0,8571	–ln(0,8571) =0,1542
22	17	1	16	0,8067	–ln(0,8067) =0,2148
29	15	1	14	0,7529	–ln(0,7529) =0,2838
35	14	1	13	0,6992	–ln(0,6992) =0,3578
37	13	1	12	0,6454	–ln(0,6454) =0,4379
39	12	1	11	0,5916	–ln(0,5916) =0,5249
40	11	1	10	0,5378	–ln(0,5378) =0,6203
52	9	1	8	0,4781	–ln(0,4781) =0,7379
54	8	1	7	0,4183	–ln(0,4183) =0,8716
60	7	1	6	0,3585	–ln(0,3585) =1,0258
80	6	1	5	0,2988	–ln(0,2988) =1,2080
83	3	1	2	0,1992	–ln(0,1992) =1,6134
84	2	1	1	0,0996	–ln(0,0996) =2,3066

- Comandos no R:

```
> Lambda.km = -log(sobrev.km$surv)
> time.km = sobrev.km$time
> S.km = sobrev.km$surv
> cbind(time.km, S.km, Lambda.km)
      time.km       S.km  Lambda.km
 [1,]      16 0.95238095 0.04879016
 [2,]      18 0.90476190 0.10008346
 [3,]      21 0.85714286 0.15415068
 [4,]      22 0.80672269 0.21477530
 [5,]      25 0.80672269 0.21477530
 [6,]      29 0.75294118 0.28376817
 [7,]      35 0.69915966 0.35787615
 [8,]      37 0.64537815 0.43791885
 [9,]      39 0.59159664 0.52493023
[10,]      40 0.53781513 0.62024041
[11,]      50 0.53781513 0.62024041
[12,]      52 0.47805789 0.73802345
[13,]      54 0.41830065 0.87155484
[14,]      60 0.35854342 1.02570552
[15,]      80 0.29878618 1.20802707
[16,]      81 0.29878618 1.20802707
[17,]      83 0.19919079 1.61349218
[18,]      84 0.09959539 2.30663936
[19,]      85 0.09959539 2.30663936
```

- Gráfico da função de risco acumulado
  - √ Estimativa por Kaplan-Meier

```
> plot(sobrev.km, conf.int = F, xlab = "Tempo", ylab = "S(t)", fun = "cumhaz")
```



$$\hat{\Lambda}_{KM}(21) = -\ln(\hat{S}_{KM}(21)) = -\ln(0,8571) = 0,1542$$

Estimador de Nelson-Aalen

$$\hat{\Lambda}_{NA}(t) = \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)}$$

√ Pode-se estimar a sobrevivência por:

$$\hat{S}_{NA}(t) = \exp\left(-\hat{\Lambda}_{NA}(t)\right)$$

√ Indicado para amostras muito pequenas

√ Para amostras suficientemente grandes as estimativas de K-M e N-A para a função de sobrevivência são bem próximas

t_j	*R(t)*	**Δ*N(t)***	$\hat{S}_{KM}(t)$	$\hat{\Lambda}_{KM}$	$\hat{\Lambda}_{NA}$
16	21	1	0,9524	0,0488	(1/21)=0,0476
18	20	1	0,9048	0,1001	(0,0476+1/20)=0,0976
21	19	1	0,8571	0,1542	(0,0976+1/19)=0,1503
22	17	1	0,8067	0,2148	(0,1503+1/17)=0,2091
29	15	1	0,7529	0,2838	(0,2091+1/15)=0,2757
35	14	1	0,6992	0,3578	(0,2757+1/14)=0,3472
37	13	1	0,6454	0,4379	(0,3472+1/13)=0,4241
39	12	1	0,5916	0,5249	(0,4241+1/12)=0,5074
40	11	1	0,5378	0,6203	(0,5074+1/11)=0,5983
52	9	1	0,4781	0,7379	(0,5953+1/9)=0,7094
54	8	1	0,4183	0,8716	(0,7094+1/8)=0,8344
60	7	1	0,3585	1,0258	(0,8344+1/7)=0,9773
80	6	1	0,2988	1,2080	(0,9773+1/6)=1,1440
83	3	1	0,1992	1,6134	(1,1440+1/3)=1,4773
84	2	1	0,0996	2,3066	(1,4773+1/2)=1,9773

- Estimativa Risco Acumulado (Nelson-Aalen)

```
> cumhaz.na<- cumsum(sobrev.km$n.event/sobrev.km$n.risk)
> cumhaz.tab<-cbind(sobrev.km$time,sobrev.km$n.event,sobrev.km$n.risk, cumhaz.na)
> colnames(cumhaz.tab)<- c("time", "n.event", "n.risk", "cumhaz.na")
> cumhaz.tab
      time n.event n.risk  cumhaz.na
 [1,]   16       1     21 0.04761905
 [2,]   18       1     20 0.09761905
 [3,]   21       1     19 0.15025063
 [4,]   22       1     17 0.20907416
 [5,]   25       0     16 0.20907416
 [6,]   29       1     15 0.27574082
 [7,]   35       1     14 0.34716939
 [8,]   37       1     13 0.42409247
 [9,]   39       1     12 0.50742580
[10,]   40       1     11 0.59833490
[11,]   50       0     10 0.59833490
[12,]   52       1      9 0.70944601
[13,]   54       1      8 0.83444601
[14,]   60       1      7 0.97730315
[15,]   80       1      6 1.14396982
[16,]   81       0      4 1.14396982
[17,]   83       1      3 1.47730315
[18,]   84       1      2 1.97730315
[19,]   85       0      1 1.97730315
```

t_j	$R(t)$	$\Delta N(t)$	$\hat{S}_{KM}(t)$	$\hat{\Lambda}_{NA}$	$\hat{S}_{NA}$
16	21		0,9524	0,0476	exp(–0,0476)=0,9535
18	20		0,9048	0,0976	exp(–0,0976)=0,9070
21	19		0,8571	0,1503	exp(–0,1503)=0,8604
22	17		0,8067	0,2091	exp(–0,2091)=0,8113
29	15		0,7529	0,2757	exp(–0,2757)=0,7590
35	14		0,6992	0,3472	exp(–0,3472)=0,7067
37	13		0,6454	0,4241	exp(–0,4241)=0,6544
39	12		0,5916	0,5074	exp(–0,5074)=0,6021
40	11		0,5378	0,5983	exp(–0,5983)=0,5497
52	9		0,4781	0,7094	exp(–0,7094)=0,4919
54	8		0,4183	0,8344	exp(–0,8344)=0,4341
60	7		0,3585	0,9773	exp(–0,9773)=0,3763
80	6		0,2988	1,1440	exp(–1,1440)=0,3185
83	3		0,1992	1,4773	exp(–1,4773)=0,2283
84	2		0,0996	1,9773	exp(–1,9773)=0,1384

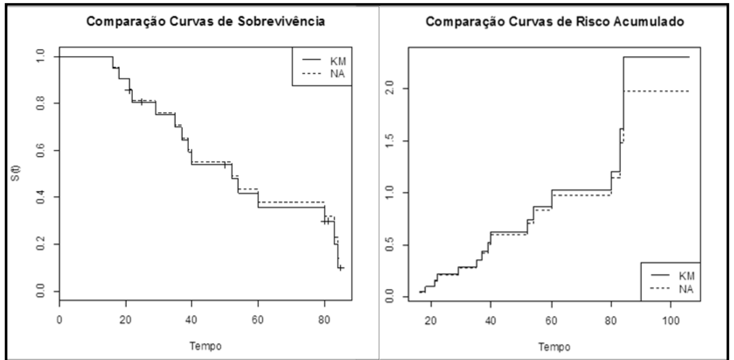
- Estimativa Nelson-Aalen para a sobrevivência:

```
> sobrev.na <- survfit(coxph(Surv(tempo, status)~1, data = ipec90))
> summary(sobrev.na)
```

```
 time n.risk n.event survival
   16     21       1    0.953
   18     20       1    0.907
   21     19       1    0.860
   22     17       1    0.811
   29     15       1    0.759
   35     14       1    0.707
   37     13       1    0.654
   39     12       1    0.602
   40     11       1    0.550
   52      9       1    0.492
   54      8       1    0.434
   60      7       1    0.376
   80      6       1    0.319
   83      3       1    0.228
   84      2       1    0.138
```

$$\hat{\Lambda}_{NA}(18) = \frac{1}{21} + \frac{1}{20} = 0,0976$$

$$\hat{S}_{NA}(18) = \exp(-\hat{\Lambda}_{NA}(t)) = \exp(-0,0976) = 0,9070$$



$$\hat{S}_{NA}(t) \geq \hat{S}_{KM}(t) \Rightarrow \hat{\Lambda}_{NA} \leq \hat{\Lambda}_{KM}, \forall t$$

Propriedades do Estimador de Kaplan-Meier

- É estimador de máxima verossimilhança de S(t)
  - √ É não viciado para amostras grandes
  - √ É fracamente consistente
  - √ Converge assintoticamente para um processo gaussiano

Intervalos de Confiança

- Variância do estimador Kaplan-Meier
  - √ Estimador de Greenwood:

$$\widehat{\text{Var}}\left[\hat{S}_{KM}(t)\right] = \left[\hat{S}_{KM}(t)\right]^2 \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]}$$

  - √ É uma expressão assintótica

- Intervalo de confiança linear para S(t):
  - √ O estimador KM, para t fixo, tem distribuição assintótica normal
  - √ Intervalo aproximado de 100 (1 – α)% de confiança para S(t)

$$\hat{S}(t) \pm z_{\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{S}_{KM}(t))}$$

  - √ Este intervalo pode levar a limites incompatíveis com a definição de sobrevivência:

    (permite valores negativos e maiores do que 1)

Relação entre Estimadores de Kaplan-Meier

$$\hat{\Lambda}_{KM}(t) = \ln(\hat{S}_{KM}(t))$$

- Variâncias dos estimadores de Kaplan-Meier da sobrevivência e do risco acumulado:

$$\widehat{\text{Var}}\left[\hat{\Lambda}_{KM}(t)\right] = \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]}$$

$$\widehat{\text{Var}}\left[\hat{S}_{KM}(t)\right] = \left[\hat{S}_{KM}(t)\right]^2 \widehat{\text{Var}}\left[\hat{\Lambda}_{KM}(t)\right]$$

- Obtenção do erro padrão $\widehat{\text{ep}}\left[\hat{\Lambda}_{KM}(t)\right] = \sqrt{\widehat{\text{Var}}\left[\hat{\Lambda}_{KM}(t)\right]}$ no R
  - √ **`survfit.object$std.err`**

Exemplo

- Sobrevivência ao diagnóstico de Aids
  - √ Pacientes com tempo de observação menor que 90 dias
  - √ Construir intervalos com 95% de confiança para S(18) e S(84).

- Intervalo de confiança linear para S(18):
  - √ Estimativa Kaplan-Meier: $\hat{S}_{KM}(18) = 0,9048$
  - √ Erro padrão da estimativa:

$$\widehat{\text{Var}}[\hat{S}_{KM}(18)] = \left[\hat{S}_{KM}(18)\right]^2 \sum_{t_j \leq 18} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]}$$
$$= (0,9048)^2 \left[\frac{1}{21(21-1)} + \frac{1}{20(20-1)}\right]$$
$$= 0,0041036$$
$$\widehat{\text{ep}}[\hat{S}_{KM}(18)] = \sqrt{\widehat{\text{Var}}[\hat{S}_{KM}(18)]} = \sqrt{0,0041036}$$
$$= 0,0641$$

  - √ Intervalo com 95% de confiança:

$$\hat{S}_{KM}(18) \pm z_{\alpha/2} \times \widehat{\text{ep}}[\hat{S}_{KM}(18)] = 0,9048 \pm (1,96)(0,0641)$$
$$\boxed{[0,7792; 1,0300]} = [0,7792; 1,0000]$$

- Intervalo de confiança linear para S(84):
  - √ Estimativa Kaplan-Meier: $\hat{S}_{KM}(84) = 0,0996$
  - √ Erro padrão da estimativa:

$$\widehat{\text{Var}}[\hat{S}_{KM}(84)] = \left[\hat{S}_{KM}(84)\right]^2 \sum_{t_j \leq 84} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]}$$
$$= (0,0996)^2 \left[\frac{1}{21(21-1)} + \cdots + \frac{1}{2(2-1)}\right]$$
$$= 0,0079404$$
$$\widehat{\text{ep}}[\hat{S}_{KM}(84)] = \sqrt{\widehat{\text{Var}}[\hat{S}_{KM}(84)]} = \sqrt{0,0079404}$$
$$= 0,0891$$

  - √ Intervalo com 95% de confiança:

$$\hat{S}_{KM}(84) \pm z_{\alpha/2}\widehat{\text{ep}}[\hat{S}_{KM}(84)] = 0,0996 \pm (1,96)(0,0891)$$
$$\boxed{[-0,0751; 0,2743]} = [0,0000; 0,2743]$$

- Construção do intervalo de confiança linear – R:
  - √ Opção survfit: **`conf.type = “plain”`**

```
> sobrev.km<-survfit( formula = Surv(tempo, status)~1, data = ipec90,
+ conf.type="plain")
> summary(sobrev.km)
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec90, conf.type =
"plain")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     21       1   0.9524  0.0465       0.8613        1.000
   18     20       1   0.9048  0.0641       0.7792        1.000
   21     19       1   0.8571  0.0764       0.7075        1.000
   22     17       1   0.8067  0.0869       0.6363        0.977
   29     15       1   0.7529  0.0963       0.5641        0.942
   35     14       1   0.6992  0.1034       0.4965        0.902
   37     13       1   0.6454  0.1085       0.4327        0.858
   39     12       1   0.5916  0.1120       0.3720        0.811
   40     11       1   0.5378  0.1140       0.3143        0.761
   52      9       1   0.4781  0.1160       0.2508        0.705
   54      8       1   0.4183  0.1158       0.1913        0.645
   60      7       1   0.3585  0.1137       0.1358        0.581
   80      6       1   0.2988  0.1093       0.0845        0.513
   83      3       1   0.1992  0.1092       0.0000        0.413
   84      2       1   0.0996  0.0891       0.0000        0.274
```


- Comandos para estimativas individuais:
  - √ Opção survfit: **`conf.type = “plain”`**
  - √ Erro padrão da estimativa da sobrevivência:
    - summary.object$std.err[summary.object$time == *t*]

```
> sobrev.km.summary <- summary(sobrev.km)
> sobrev.km.summary$std.err[sobrev.km.summary$time==84]
[1] 0.08910877
```

  - √ Limite inferior do intervalo de confiança
    - survfit.object$lower[survfit.object$time == *t*]

```
> sobrev.km$lower[sobrev.km$time==84]
[1] 0
```

  - √ Limite superior do intervalo de confiança
    - survfit.object$upper[survfit.object$time == *t*]

```
> sobrev.km$upper[sobrev.km$time==84]
[1] 0.2742454
```

- Cálculo do erro padrão das estimativas
  - √ Erro padrão de log(S(t)):

```
> tabela.var = rbind(
+ c(18, sobrev.km$surv[sobrev.km$time==18],
+ sobrev.km$std.err[sobrev.km$time==18])
+ ,
+ c(84, sobrev.km$surv[sobrev.km$time==84],
+ sobrev.km$std.err[sobrev.km$time==84])
+ )
> colnames(tabela.var) = c("Tempo", "Sobrevivencia", "ep(ln S(t))" )
> tabela.var
      Tempo Sobrevivencia ep(ln S(t))
[1,]    18    0.90476190  0.07079923
[2,]    84    0.09959539  0.89470770
```

  - √ Erro padrão de S(t)

$$\widehat{\text{ep}}[\hat{S}_{KM}(18)] = \hat{S}_{KM}(18) \times \widehat{\text{ep}}\left[\hat{\lambda}_{KM}(18)\right] = (0,9048)(0,0708)$$
$$= 0,0641$$
$$\widehat{\text{ep}}[\hat{S}_{KM}(84)] = \hat{S}_{KM}(84) \times \widehat{\text{ep}}\left[\hat{\Lambda}_{KM}(84)\right] = (0,0996)(0,8947)$$
$$= 0,0891$$

√ Estimativa do intervalo de confiança exponencial de Grenwood para S(t):

				Limites			
				$\hat{S}_{KM}(t)$		Truncamentos	
t_j	$\hat{S}_{KM}(t)$	$\ln(\hat{\Lambda}_{KM}(t))$	Erro padrão(*)	Inferior	Superior	Inferior	Superior
16	0,9524	-3,0202	0,0465	0,8613	1,0435	0,8613	1
18	0,9048	-2,3018	0,0641	0,7792	1,0304	0,7792	1
21	0,8571	-1,8698	0,0764	0,7074	1,0068	0,7075	1
22	0,8067	-1,5382	0,0869	0,6364	0,9770		
29	0,7529	-1,2596	0,0963	0,5642	0,9416		
35	0,6992	-1,0276	0,1034	0,4965	0,9019		
37	0,6454	-0,8257	0,1085	0,4327	0,8581		
39	0,5916	-0,6445	0,112	0,3721	0,8111		
40	0,5378	-0,4776	0,114	0,3144	0,7612		
52	0,4781	-0,3038	0,116	0,2507	0,7055		
54	0,4183	-0,1375	0,1158	0,1913	0,6453		
60	0,3585	0,0254	0,1137	0,1357	0,5813		
80	0,2988	0,189	0,1093	0,0846	0,5130		
83	0,1992	0,4784	0,1092	-0,0148	0,4132	0	0,4132
84	0,0996	0,8358	0,0891	-0,0750	0,2742	0	0,2742

(*) Erro padrão de Greenwood para $\hat{S}_{KM}(t)$

$$\hat{S}_{KM}(t) \pm z_{\alpha/2}\, \widehat{\text{ep}}\left[\hat{S}_{KM}(t)\right]$$

√ Construção da tabela em R:

```
> alpha = 0.05
> casas = 4
> z = qnorm(1-alpha/2)
> tj = sobrev.km$time
> S.km = round(sobrev.km$surv, casas)
>
> # Tabela 4.3b Estimativa IC tradicional de Greenwood
>
> ep.S.km <- round(sobrev.km$std.err * sobrev.km$surv, casas)
> inf <- round(S.km - z*ep.S.km, casas)
> sup <- round(S.km + z*ep.S.km, casas)
> inf.trunc <- round(sobrev.km$lower, casas)
> sup.trunc <- round(sobrev.km$upper,casas)
> tab.43b<-cbind(tj, S.km, ep.S.km, inf, sup, inf.trunc, sup.trunc)
> colnames(tab.43b)<- c("tj", "Skm(t)", "ep[Skm(t)]", "inf", "sup",
+ "inf final", "sup final")
> tab.43b
      tj Skm(t) ep[Skm(t)]    inf    sup inf final sup final
 [1,] 16 0.9524     0.0465 0.8613 1.0435    0.8613    1.0000
 [2,] 18 0.9048     0.0641 0.7792 1.0304    0.7792    1.0000
 [3,] 21 0.8571     0.0764 0.7075 1.0068    0.7075    1.0000
 [4,] 22 0.8067     0.0869 0.6363 0.9771    0.6363    0.9771
 [5,] 25 0.8067     0.0869 0.6363 0.9771    0.6363    0.9771
 [6,] 29 0.7529     0.0963 0.5641 0.9418    0.5641    0.9418
 [7,] 35 0.6992     0.1034 0.4965 0.9018    0.4965    0.9018
 [8,] 37 0.6454     0.1085 0.4327 0.8581    0.4327    0.8581
 [9,] 39 0.5916     0.1120 0.3720 0.8112    0.3720    0.8112
[10,] 40 0.5378     0.1140 0.3143 0.7613    0.3143    0.7613
[11,] 50 0.5378     0.1140 0.3143 0.7613    0.3143    0.7613
```


- Procedimento de construção dos intervalos de confiança lineares levou a resultados incoerentes com a definição de probabilidade
- Solução:
  - √ Construir intervalos assimétricos para S(t), a partir de transformações de intervalos simétricos

- Alternativa para contornar incompatibilidade em limites:
  - √ Construir intervalo simétrico para ln(Λ(t))
  - √ Construir intervalo para S(t) através da transformação U(t) = ln(Λ(t)) = ln(–ln(S(t))
  - √ Intervalo será assimétrico, porém positivo e menor ou igual a 1
- É considerado mais preciso:
  - √ $\ln(\hat{\Lambda}_{KM}(t))$ tem distribuição mais próxima da normal que $\hat{S}_{KM}(t)$

- Transformação U(t) = ln(Λ(t)) = ln[–ln(S(t))]:
  - √ Estimador da variância assintótica de ln(Λ(t))

$$\widehat{\mathrm{Var}}[\ln(\hat{\Lambda}_{KM}(t))] = \frac{1}{[\ln(\hat{S}_{KM}(t))]^2} \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)(R(t_j) - \Delta N(t_j))}$$

$$\widehat{\mathrm{Var}}\left[\ln(\hat{\Lambda}_{KM}(t))\right] = \frac{1}{\left[\ln(\hat{S}_{KM}(t))\right]^2} \widehat{\mathrm{Var}}\left[\hat{\Lambda}_{KM}(t)\right]$$

  - √ Erro padrão do estimador Kaplan-Meier de ln(Λ(t)):

$$\widehat{\mathrm{ep}}\left[\ln(\hat{\Lambda}_{KM}(t))\right] = \sqrt{\widehat{\mathrm{Var}}\left[\ln(\hat{\Lambda}_{KM}(t))\right]}$$

- √ Intervalo de (1 – α)100% de confiança para ln(Λ(t))

$$l_{inf} = \ln\left(\hat{\Lambda}_{KM}(t)\right) - z_{\alpha/2}\, \widehat{\mathrm{ep}}\left[\ln(\hat{\Lambda}_{KM}(t))\right]$$
$$l_{sup} = \ln\left(\hat{\Lambda}_{KM}(t)\right) + z_{\alpha/2}\, \widehat{\mathrm{ep}}\left[\ln(\hat{\Lambda}_{KM}(t))\right]$$

- √ Intervalo de confiança exponencial de Greenwood para S(t):

$$\text{limite inferior} = \exp\{-\exp\{l_{sup}\}\}$$
$$\text{limite superior} = \exp\{-\exp\{l_{inf}\}\}$$

- √ Construção no R:
  - Opção survfit: **`conf.type = “log-log”`**

- Intervalo exponencial de confiança para S(18):
  - √ Estimativa Kaplan-Meier: $\hat{S}_{KM}(18) = 0,9048$
  - √ Erro padrão da estimativa:

$$\widehat{\mathrm{Var}}[\ln(\hat{\Lambda}_{KM}(18))] = \frac{1}{\left[\ln(\hat{S}_{KM}(18))\right]^2} \sum_{t_j \leq 18} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]}$$
$$= \frac{1}{\ln(0,9048)^2}\left[\frac{1}{21(21-1)} + \frac{1}{20(20-1)}\right]$$
$$= 0,50084$$
$$\widehat{\mathrm{ep}}[\ln(\hat{\Lambda}_{KM}(18))] = \sqrt{\widehat{\mathrm{Var}}[\ln(\hat{\Lambda}_{KM}(18))]} = \sqrt{0,50084}$$
$$= 0,7077$$

  - √ Intervalo simétrico para ln(Λ(t)) :

$$l_{inf} = \ln(-\ln(0,9048)) - (1,96)(0,7077) = -3,6893$$
$$l_{sup} = \ln(-\ln(0,9048)) + (1,96)(0,7077) = -0,9155$$

√ Intervalo de confiança exponencial de Greenwood para S(t):

$$\begin{aligned} \text{limite inferior} &= \exp(-\exp(l_{sup})) \\ &= \exp(-\exp(-0,9155)) \\ &= 0,6701 \\ \text{limite superior} &= \exp(-\exp(l_{inf})) \\ &= \exp(-\exp(-3,6893)) \\ &= 0,9753 \end{aligned}$$

$$[0,6700; 0,9753]$$

- Expressão alternativa para obtenção do intervalo aproximado de (1 – α)100% de confiança para S(t):
  - √ (intervalo exponencial de Greenwood):

$$\left[\hat{S}_{KM}(t)\right]^{\exp\left\{\pm z_{\alpha/2}\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(t))\right]\right\}}$$

- Intervalo exponencial de confiança para S(18):
  - √ Estimativa Kaplan-Meier: $\hat{S}_{KM}(18) = 0,9048$
  - √ Erro padrão da estimativa: $\widehat{\text{ep}}[\ln(\hat{\Lambda}_{KM}(18))] = 0,7077$
  - √ Intervalo exponencial de confiança para S(t):

$$\begin{aligned} \text{limite inferior} &= \left[\hat{S}_{KM}(18)\right]^{\exp\left\{1,96\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(18))\right]\right\}} \\ &= [0,9048]^{\exp\{(1,96)(0,7077)\}} \\ &= 0,6700 \\ \text{limite superior} &= \left[\hat{S}_{KM}(18)\right]^{\exp\left\{-1,96\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(18))\right]\right\}} \\ &= [0,9048]^{\exp\{(-1,96)(0,7077)\}} \\ &= 0,9753 \end{aligned}$$

- Construção do intervalo de confiança no R:
  - √ Opção survfit: **`conf.type = “log-log”`**

```
> sobrev.km<-survfit( formula = Surv(tempo, status)~1, data = ipec90,
+ conf.type="log-log")
> summary(sobrev.km)
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec90, conf.type = "log-
log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     21       1   0.9524  0.0465       0.7072        0.993
   18     20       1   0.9048  0.0641       0.6700        0.975
   21     19       1   0.8571  0.0764       0.6197        0.952
   22     17       1   0.8067  0.0869       0.5631        0.923
   29     15       1   0.7529  0.0963       0.5032        0.889
   35     14       1   0.6992  0.1034       0.4474        0.853
   37     13       1   0.6454  0.1085       0.3947        0.814
   39     12       1   0.5916  0.1120       0.3449        0.772
   40     11       1   0.5378  0.1140       0.2976        0.728
   52      9       1   0.4781  0.1160       0.2452        0.679
   54      8       1   0.4183  0.1158       0.1970        0.627
   60      7       1   0.3585  0.1137       0.1526        0.571
   80      6       1   0.2988  0.1093       0.1123        0.513
   83      3       1   0.1992  0.1092       0.0433        0.436
   84      2       1   0.0996  0.0891       0.0072        0.340
```

- Comandos para estimativas individuais:
  - √ Opção survfit: **`conf.type = "log-log"`**
  - √ Limite inferior do intervalo de confiança
    - survfit.object$lower[survfit.object$time == *t*]

```
> sobrev.km$lower[sobrev.km$time==18]
[1] 0.6700459
```

  - √ Limite superior do intervalo de confiança
    - survfit.object$upper[survfit.object$time == *t*]

```
> sobrev.km$upper[sobrev.km$time==18]
[1] 0.9752941
```


- Cálculo do erro padrão das estimativas
  - √ Erro padrão de log(S(t)):

```
> tabela.var = rbind(
+ c(18, sobrev.km$surv[sobrev.km$time==18],
+ sobrev.km$std.err[sobrev.km$time==18])
+ ,
+ c(84, sobrev.km$surv[sobrev.km$time==84],
+ sobrev.km$std.err[sobrev.km$time==84])
+ )
> colnames(tabela.var) = c("Tempo", "Sobrevivencia", "ep(ln S(t))" )
> tabela.var
     Tempo Sobrevivencia ep(ln S(t))
[1,]    18    0.90476190  0.07079923
[2,]    84    0.09959539  0.89470770
```

  - √ Erro padrão de ln(Λ(t))

$$\widehat{\text{ep}}[\ln(\hat{\Lambda}_{KM}(18))] = \frac{\widehat{\text{ep}}\left[\hat{\lambda}_{KM}(18)\right]}{-\ln(\hat{S}_{KM}(18))} = \frac{(0,0708)}{-\ln(0,9048)}$$
$$= 0,70740188$$
$$\widehat{\text{ep}}[\ln(\hat{\Lambda}_{KM}(84))] = \frac{\widehat{\text{ep}}\left[\hat{\Lambda}_{KM}(84)\right]}{-\ln(\hat{S}_{KM}(84))} = \frac{0,8947}{-\ln(0,0996)}$$
$$= 0,3879$$

- Intervalo exponencial de confiança para S(84):
  - √ Estimativa Kaplan-Meier: $\hat{S}_{KM}(18) = 0,0996$
  - √ Erro padrão da estimativa: $\widehat{\text{ep}}[\ln(\hat{\Lambda}_{KM}(18))] = 0,3879$
  - √ Intervalo exponencial de confiança para S(t):

$$\text{limite inferior} = \left[\hat{S}_{KM}(84)\right]^{\exp\left\{1,96\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(84))\right]\right\}}$$
$$= [0,0996]^{\exp\{(1,96)(0,3879)\}}$$
$$= 0,0072$$
$$\text{limite superior} = \left[\hat{S}_{KM}(84)\right]^{\exp\left\{-1,96\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(84))\right]\right\}}$$
$$= [0,0996]^{\exp\{(-1,96)(0,3879)\}}$$
$$= 0,3401$$

√ Estimativa do intervalo de confiança exponencial de Grenwood para S(t):

				Limites			
				$\ln(\hat{\Lambda}_{KM}(t))$		$\hat{S}_{KM}(t)$	
t_j	$\hat{S}_{KM}(t)$	$\ln(\hat{\Lambda}_{KM}(t))$	Erro padrão(*)	Inferior	Superior	Inferior	Superior
16	0,9524	-3,0202	1,0001	-4,9804	-1,06	0,7072	0,9932
18	0,9048	-2,3018	0,7074	-3,6883	-0,9153	0,6701	0,9753
21	0,8571	-1,8698	0,5779	-3,0025	-0,7371	0,6197	0,9516
22	0,8067	-1,5382	0,5018	-2,5217	-0,5547	0,5631	0,9228
29	0,7529	-1,2596	0,4509	-2,1433	-0,3759	0,5032	0,8893
35	0,6992	-1,0276	0,4132	-1,8375	-0,2177	0,4474	0,8528
37	0,6454	-0,8257	0,384	-1,5783	-0,0731	0,3947	0,8136
39	0,5916	-0,6445	0,3607	-1,3515	0,0625	0,3449	0,7719
40	0,5378	-0,4776	0,3418	-1,1475	0,1923	0,2976	0,728
52	0,4781	-0,3038	0,3287	-0,948	0,3404	0,2452	0,6787
54	0,4183	-0,1375	0,3177	-0,7602	0,4852	0,197	0,6265
60	0,3585	0,0254	0,3091	-0,5804	0,6312	0,1526	0,5714
80	0,2988	0,189	0,3028	-0,4045	0,7825	0,1123	0,5131
83	0,1992	0,4784	0,3397	-0,1874	1,1442	0,0433	0,4364
84	0,0996	0,8358	0,3879	0,0755	1,5961	0,0072	0,3401

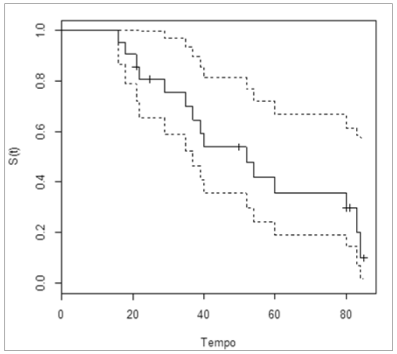
(*) Erro padrão de Greenwood para $\ln(\hat{\Lambda}_{KM}(t))$

$$l_{inf} = \ln\left(\hat{\Lambda}_{KM}(t)\right) - z_{\alpha/2}\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(t))\right]$$
$$l_{sup} = \ln\left(\hat{\Lambda}_{KM}(t)\right) + z_{\alpha/2}\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{KM}(t))\right]$$
$$\text{limite inferior} = \exp\{-\exp\{l_{sup}\}\}$$
$$\text{limite superior} = \exp\{-\exp\{l_{inf}\}\}$$

√ Construção da tabela em R:

```
> # Tabela 4.4 Estimativa IC exponencial de Greenwood
>
> alpha = 0.05
> casas = 4
> z = qnorm(1-alpha/2)
> tj = sobrev.km$time
> S.km = round(sobrev.km$surv, casas)
> ln.Lambda = round(log(-log(sobrev.km$surv)), casas)
> ep.lnLambda = round(sobrev.km$std.err/(-log(sobrev.km$surv)), casas)
> li = round(ln.Lambda - z*ep.lnLambda, casas)
> ls = round(ln.Lambda + z*ep.lnLambda, casas)
>
> linf = round(exp(-exp(ls)), casas)
> lsup = round(exp(-exp(li)), casas)
>
> tab.44<-cbind(tj, S.km, ln.Lambda, ep.lnLambda, li, ls, linf, lsup)
> colnames(tab.44)<- c("tj", "Skm(t)", "ln(Lambda(t))", "ep[ln(Lambda(t))]",
+ "li", "ls", "inf", "sup")
> tab.44
     tj Skm(t) ln(Lambda(t)) ep[ln(Lambda(t))]      li      ls    inf    sup
 [1,] 16 0.9524       -3.0202            1.0001 -4.9804 -1.0600 0.7072 0.9932
 [2,] 18 0.9048       -2.3018            0.7074 -3.6883 -0.9153 0.6701 0.9753
 [3,] 21 0.8571       -1.8698            0.5779 -3.0025 -0.7371 0.6197 0.9516
 [4,] 22 0.8067       -1.5382            0.5018 -2.5217 -0.5547 0.5631 0.9228
 [5,] 25 0.8067       -1.5382            0.5018 -2.5217 -0.5547 0.5631 0.9228
 [6,] 29 0.7529       -1.2596            0.4509 -2.1433 -0.3759 0.5032 0.8893
 [7,] 35 0.6992       -1.0276            0.4132 -1.8375 -0.2177 0.4474 0.8528
 [8,] 37 0.6454       -0.8257            0.3840 -1.5783 -0.0731 0.3947 0.8136
 [9,] 39 0.5916       -0.6445            0.3607 -1.3515  0.0625 0.3449 0.7719
```


√ Curva de sobrevivência estimada pelo Kaplan-Meier com intervalo de confiança de 95%



√ Estimativas do intervalo de confiança exponencial de Greenwood

Comentários

- Intervalo exponencial de confiança de Greenwood
  - √ Garantia de os limites do intervalo estarem entre 0 e 1
  - √ Intervalo comporta-se bem para amostras pequenas ($n \geq 25$) com até 50% de observações censuradas

    (Hosmer e Lemesehow, 1999)

- Transformação $U(t) = \Lambda(t) = -\ln(S(t))$:
  - √ Estimador da variância assintótica de $\Lambda(t)$

$$\widehat{\mathrm{Var}}[\hat{\Lambda}_{KM}(t)] = \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)(R(t_j) - \Delta N(t_j))}$$

$$\widehat{\mathrm{Var}}\left[\ln(\hat{\Lambda}_{KM}(t))\right] = \frac{1}{\left[\ln(\hat{S}_{KM}(t))\right]^2} \widehat{\mathrm{Var}}\left[\hat{\Lambda}_{KM}(t)\right]$$

  - √ Erro padrão do estimador Kaplan-Meier de $\Lambda(t)$:

$$\widehat{\mathrm{ep}}\left[\hat{\Lambda}_{KM}(t)\right] = \sqrt{\widehat{\mathrm{Var}}\left[\hat{\Lambda}_{KM}(t)\right]}$$

√ Intervalo de $(1 - \alpha)100\%$ de confiança para $\Lambda(t)$:

$$l_{inf} = -\ln\left(\hat{S}_{KM}(t)\right) - z_{\alpha/2}\,\widehat{\text{ep}}\left[\hat{\Lambda}_{KM}(t)\right]$$
$$l_{sup} = -\ln\left(\hat{S}_{KM}(t)\right) + z_{\alpha/2}\,\widehat{\text{ep}}\left[\hat{\Lambda}_{KM}(t)\right]$$

√ Intervalo de confiança assimétrico para S(t):

$$\text{limite inferior} = \exp\{-l_{sup}\}$$
$$\text{limite superior} = \exp\{-l_{inf}\}$$

√ Pode resultar limite superior maior que 1!

√ Construção no R:

- Opção survfit: **`conf.type = “log”`** (default)

- Intervalo de confiança para S(18):

√ Estimativa Kaplan-Meier: $\hat{S}_{KM}(18) = 0,9048$

√ Erro padrão da estimativa:

$$\begin{aligned}\widehat{\text{Var}}[\hat{\Lambda}_{KM}(18)] &= \sum_{t_j \le 18} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]} \\ &= \frac{1}{21(21-1)} + \frac{1}{20(20-1)} \\ &= 0,0501253 \end{aligned}$$

$$\begin{aligned}\widehat{\text{ep}}[\hat{\Lambda}_{KM}(18)] &= \sqrt{\widehat{\text{Var}}[\hat{\Lambda}_{KM}(18)]} = \sqrt{0,00501253} \\ &= 0,0708\end{aligned}$$

√ Intervalo simétrico para $\Lambda(t)$:

$$l_{inf} = -\ln(0,9048) - (1,96)(0,0708) = -0,0387$$
$$l_{sup} = -\ln(0,9048) + (1,96)(0,0708) = 0,2388$$

√ Intervalo de confiança para S(t):

- U(t) = – log(S(t))

$$\begin{aligned}\text{limite inferior} &= \exp\{-l_{sup}\} \\ &= \exp\{-0,2388\} \\ &= 0,7875 \\ \text{limite superior} &= \exp\{-l_{inf}\} \\ &= \exp\{-0,0387\} \\ &= 1,0394\end{aligned}$$

$$[0,7875; 1,0000]$$

- Construção do intervalo de confiança no R:

√ Opção survfit: **`conf.type = “log”`**

```
> sobrev.km<-survfit( formula = Surv(tempo, status)~1, data = ipec90,
+ conf.type="log")
> summary(sobrev.km)
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec90, conf.type =
"log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     21       1   0.9524  0.0465       0.8655        1.000
   18     20       1   0.9048  0.0641       0.7875        1.000
   21     19       1   0.8571  0.0764       0.7198        1.000
   22     17       1   0.8067  0.0869       0.6531        0.996
   29     15       1   0.7529  0.0963       0.5859        0.968
   35     14       1   0.6992  0.1034       0.5232        0.934
   37     13       1   0.6454  0.1085       0.4642        0.897
   39     12       1   0.5916  0.1120       0.4082        0.857
   40     11       1   0.5378  0.1140       0.3550        0.815
   52      9       1   0.4781  0.1160       0.2972        0.769
   54      8       1   0.4183  0.1158       0.2431        0.720
   60      7       1   0.3585  0.1137       0.1926        0.667
   80      6       1   0.2988  0.1093       0.1459        0.612
   83      3       1   0.1992  0.1092       0.0680        0.583
   84      2       1   0.0996  0.0891       0.0172        0.575
```

- Comandos para estimativas individuais:
  - √ Opção survfit: **`conf.type = "log"`**
  - √ Limite inferior do intervalo de confiança
    - – survfit.object$lower[survfit.object$time == *t*]

```
> sobrev.km$lower[sobrev.km$time==22]
[1] 0.6531242
```

  - √ Limite superior do intervalo de confiança
    - – survfit.object$upper[survfit.object$time == *t*]

```
> sobrev.km$upper[sobrev.km$time==22]
[1] 0.9964437
```


- Cálculo do erro padrão das estimativas
  - √ Erro padrão de log(S(t)):

```
> tabela.var = rbind(
+ c(18, sobrev.km$surv[sobrev.km$time==18],
+ sobrev.km$std.err[sobrev.km$time==18])
+ ,
+ c(22, sobrev.km$surv[sobrev.km$time==22],
+ sobrev.km$std.err[sobrev.km$time==22])
+ )
> colnames(tabela.var) = c("Tempo", "Sobrevivencia", "ep(ln S(t))" )
> tabela.var
     Tempo Sobrevivencia ep(ln S(t))
[1,]    18     0.9047619  0.07079923
[2,]    22     0.8067227  0.10776353
```

  - √ Erro padrão de S(t)

$$\widehat{\text{ep}}[\hat{\Lambda}_{KM}(18)] = 0,70740188$$

$$\widehat{\text{ep}}[\hat{\Lambda}_{KM}(22)] = 0,10776353$$

- Intervalo de confiança para S(22):
  - √ Estimativa Kaplan-Meier: $\hat{S}_{KM}(22) = 0,8067$
  - √ Erro padrão da estimativa: $\widehat{\text{ep}}[\hat{\Lambda}_{KM}(18)] = 0,1078$
  - √ Intervalo de 95% de confiança para Λ(t):

$$l_{inf} = -\ln(0,8067) - (1,96)(0,1078) = 0,00356$$
$$l_{sup} = -\ln(0,8067) + (1,96)(0,1078) = 0,42599$$

  - √ Intervalo de confiança assimétrico para S(t):

$$\begin{aligned} \text{limite inferior} &= \exp\{-l_{sup}\} = \exp\{-0,42599\} \\ &= 0,6531 \\ \text{limite superior} &= \exp\{-l_{inf}\} = \exp\{-0,00356\} \\ &= 0,9964 \end{aligned}$$

$$[0,6531; 0,9964]$$

Estimador de Nelson-Aalen

$$\hat{\Lambda}_{NA}(t) = \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)}$$

- √ A função de sobrevivência é estimada por:

$$\hat{S}_{NA}(t) = \exp\left(-\hat{\Lambda}_{NA}(t)\right)$$

- √ Indicado para amostras muito pequenas
- √ Para amostras suficientemente grandes as estimativas de K-M e N-A para a função de sobrevivência são bem próximas

Intervalos de Confiança – Nelson-Aalen

- Variância do estimador de Nelson-Aalen:
  - √ Estimador da variância (Aalen, 1978):

$$\widehat{\text{Var}}\left[\hat{\Lambda}_{NA}(t)\right] = \sum_{t_j \le t} \frac{\Delta N(t_j)}{[R(t_j)]^2}$$

  - √ Estimador alternativo (Klein, 1991):

$$\widehat{\text{Var}}\left[\hat{\Lambda}_{NA}(t)\right] = \sum_{t_j \le t} \frac{\Delta N(t_j)[R(t_j) - \Delta N(t_j)]\Delta N(t_j)}{[R(t_j)]^3}$$

  - √ É proferível estimador proposto por Aalen
    - Menor vício

- √ Estimador da função de sobrevivência:

$$\hat{S}_{NA}(t) = \exp\left\{-\hat{\Lambda}_{NA}(t)\right\}$$

- √ Estimador da variância (Aalen e Johansen, 1978)

$$\widehat{\text{Var}}\left[\hat{\Lambda}_{NA}(t)\right] = \sum_{t_j \le t} \frac{\Delta N(t_j)}{[R(t_j)]^2}$$

- √ Estimador alternativo (Kalbfleisch e Prentice, 1980):
  - Substituindo o estimador da sobrevivência na fórmula de Greenwood

$$\widehat{\text{Var}}\left[\hat{S}_{NA}(t)\right] = \left[\hat{S}_{NA}(t)\right]^2 \sum_{t_j \le t} \frac{\Delta N(t_j)}{R(t_j)[R(t_j) - \Delta N(t_j)]}$$

- Intervalo de confiança linear para S(t):
  - √ Simétrico – sem transformação:

$$\hat{S}(t) \pm z_{\alpha/2}\widehat{\text{ep}}[\hat{S}_{NA}(t)]$$

    - Pode levar a limites incompatíveis com a definição de sobrevivência
  - √ Assimétrico – transformação U(t) = ln[–ln(S(t))]
    - Estimador da variância assintótica de ln(Λ(t))

$$\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{NA}(t))\right] = \frac{\widehat{\text{ep}}\left[\hat{\Lambda}_{NA}(t)\right]}{\left[\ln(\hat{S}_{NA}(t))\right]}$$

    - Intervalo exponencial

$$\left[\hat{S}_{NA}(t)\right]^{\exp\left\{\pm z_{\alpha/2}\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{NA}(t))\right]\right\}}$$

Exemplo

- Intervalo exponencial de confiança para S(18):
  - √ Estimativa Nelson-Aalen da sobrevivência:

$$\hat{\Lambda}_{NA}(18) = \sum_{t_j \le 18} \frac{\Delta N(t_j)}{R(t_j)} = \frac{1}{21} + \frac{1}{20} = 0,097619$$

$$\hat{S}_{NA}(18) = \exp\{-0,097619\} = 0,9070$$

  - √ Erro padrão da estimativa risco acumulado:

$$\widehat{\text{Var}}[\hat{\Lambda}_{NA}(18)] = \sum_{t_j \le 18} \frac{\Delta N(t_j)}{[R(t_j)]^2} = \frac{1}{21^2} + \frac{1}{20^2} = 0,50084$$

$$\widehat{\text{ep}}[\hat{\Lambda}_{NA}(18)] = \sqrt{0,50084} = 0,06905$$

√ Erro padrão da estimativa da sobrevivência:

$$\begin{aligned}\widehat{\text{ep}}[\hat{S}_{NA}(18)] &= \hat{S}_{NA}(18)\widehat{\text{ep}}[\hat{\Lambda}_{NA}(18)\\ &= (0,9070)(0,06905)\\ &= 0,06262\end{aligned}$$

√ Erro padrão da estimativa de ln(Λ(t)):

$$\begin{aligned}\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{NA}(18))\right] &= \frac{0,06905}{-\ln(0,9070)}\\ &= 0,70732\end{aligned}$$

√ Intervalo assimétrico para S(t):

$$\begin{aligned}l_{inf} &= \left[\hat{S}_{NA}(18)\right]^{\exp\left\{1,96\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{NA}(18))\right]\right\}}\\ &= [0,9070]^{\exp\{(1,96)(0,70732)\}}\\ &= 0,6767\end{aligned}$$

$$\begin{aligned}l_{sup} &= \left[\hat{S}_{NA}(18)\right]^{\exp\left\{-1,96\,\widehat{\text{ep}}\left[\ln(\hat{\Lambda}_{NA}(18))\right]\right\}}\\ &= [0,9070]^{\exp\{-(1,96)(0,70732)\}}\\ &= 0,9759\end{aligned}$$

- Construção do intervalo de confiança no R:
  - √ Opção survfit: **`conf.type = “log-log”`**

```
> sobrev.na <- survfit(coxph(Surv(tempo, status)~1, data = ipec90,
+    method = "breslow"), conf.type = "log-log")
> summary(sobrev.na)
Call: survfit(formula = coxph(Surv(tempo, status) ~ 1, data = ipec90,
    method = "breslow"), conf.type = "log-log")
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
16	21	1	0.953	0.0454	0.7132	0.993
18	20	1	0.907	0.0626	0.6767	0.976
21	19	1	0.860	0.0747	0.6273	0.953
22	17	1	0.811	0.0851	0.5719	0.925
29	15	1	0.759	0.0943	0.5133	0.892
35	14	1	0.707	0.1013	0.4585	0.857
37	13	1	0.654	0.1064	0.4068	0.819
39	12	1	0.602	0.1100	0.3577	0.778
40	11	1	0.550	0.1122	0.3111	0.736
52	9	1	0.492	0.1143	0.2597	0.688
54	8	1	0.434	0.1146	0.2120	0.638
60	7	1	0.376	0.1129	0.1680	0.585
80	6	1	0.319	0.1093	0.1275	0.530
83	3	1	0.228	0.1092	0.0616	0.457
84	2	1	0.138	0.0958	0.0197	0.369

- Comandos para estimativas individuais:
  - √ Opção survfit: **`conf.type = “log-log”`**
  - √ Limite inferior do intervalo de confiança
    - survfit.object$lower[survfit.object$time == *t*]

```
> sobrev.na$lower[sobrev.na$time==18]
[1] 0.6767288
```

  - √ Limite superior do intervalo de confiança
    - survfit.object$upper[survfit.object$time == *t*]

```
> sobrev.na$upper[sobrev.na$time==18]
[1] 0.9758911
```

  - √ Estimativa do risco acumulado

```
> sobrev.na$cumhaz[sobrev.na$time==18]
[1] 0.09761905
```


- Cálculo do erro padrão das estimativas
  - √ Erro padrão de log(S(18)):

```
> sobrev.na$std.err[sobrev.na$time==18]
[1] 0.06904762
```

  - √ Erro padrão de ln(Λ(18))

$$\begin{aligned}\widehat{\text{ep}}[\ln(\hat{\Lambda}_{NA}(18))] &= \frac{\widehat{\text{ep}}\left[\hat{\lambda}_{NA}(18)\right]}{-\ln(\hat{S}_{NA}(18))} = \frac{(0,06905)}{-\ln(0,9070)}\\ &= 0,70732\end{aligned}$$

- Estimação Nelson-Aalen – Alternativa:
  - √ Opção survfit: **`type = “fleming-harrington”, error = “tsiatis”, conf.type = “log-log”`**

```
> sobrev.na <- survfit(Surv(tempo, status)~1, data = ipec90,
+     type = "fleming-harrington", error = "tsiatis",
+     conf.type="log-log")
> summary(sobrev.na)
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec90, type = "fleming-
harrington",
    error = "tsiatis", conf.type = "log-log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     21       1    0.953  0.0454       0.7132        0.993
   18     20       1    0.907  0.0626       0.6767        0.976
   21     19       1    0.860  0.0747       0.6273        0.953
   22     17       1    0.811  0.0851       0.5719        0.925
   29     15       1    0.759  0.0943       0.5133        0.892
   35     14       1    0.707  0.1013       0.4585        0.857
   37     13       1    0.654  0.1064       0.4068        0.819
   39     12       1    0.602  0.1100       0.3577        0.778
   40     11       1    0.550  0.1122       0.3111        0.736
   52      9       1    0.492  0.1143       0.2597        0.688
   54      8       1    0.434  0.1146       0.2120        0.638
   60      7       1    0.376  0.1129       0.1680        0.585
```


- Estimação Nelson-Aalen – Greenwood:
  - √ Opção survfit: **`type = “fleming-harrington”, error = “greenwood”, conf.type = “log-log”`**

```
> sobrev.na <- survfit(Surv(tempo, status)~1, data = ipec90,
+     type = "fleming-harrington", error = "greenwood",
+     conf.type="log-log")
> summary(sobrev.na)
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec90, type = "fleming-
harrington",
    error = "greenwood", conf.type = "log-log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     21       1    0.953  0.0465      0.70130        0.994
   18     20       1    0.907  0.0642      0.66734        0.977
   21     19       1    0.860  0.0767      0.61860        0.954
   22     17       1    0.811  0.0874      0.56317        0.927
   29     15       1    0.759  0.0971      0.50422        0.895
   35     14       1    0.707  0.1045      0.44930        0.860
   37     13       1    0.654  0.1100      0.39751        0.823
   39     12       1    0.602  0.1140      0.34839        0.783
   40     11       1    0.550  0.1165      0.30172        0.742
   52      9       1    0.492  0.1193      0.24993        0.696
   54      8       1    0.434  0.1202      0.20207        0.647
   60      7       1    0.376  0.1193      0.15793        0.596
   80      6       1    0.319  0.1165      0.11753        0.543
```

Análise de Sobrevivência Aplicada à Saúde -- 2014

Tempo de Sobrevivência Mediano

- Tempo mediano (t_{med}):
  - √ Tempo após o qual 50% dos indivíduos estão vivos

    $S(t_{med}) = 0,50$
  - √ Devido a sua robustez, é a medida-resumo mais comum em Análise de Sobrevivência
    - – Mais usado que a média pois a distribuição do tempo de sobrevivência é pouco simétrico

Análise de Sobrevivência Aplicada à Saúde -- 2014

- Estimação do tempo mediano:
  - √ A sobrevivência é uma função em degraus

    (O tempo não é observado de forma contínua)

$$\hat{t}_{med} = \min\{t_j | \hat{S}(t_j) \leq 0,5)\}$$

  - √ Situação ideal sem censura:
    - – Tempo mediano é aquele imediatamente depois do qual 50% da amostra ainda está em observação
  - √ Como a curva de Kaplan-Meier é uma função escada, a estimativa mais adequada para a mediana é obtida por meio de uma interpolação linear

Análise de Sobrevivência Aplicada à Saúde -- 2014

Exemplo

- Sobrevivência ao diagnóstico de Aids
  - √ Pacientes com tempo de observação menor que 90 dias
  - √ Estimar o tempo mediano de sobrevida.
  - √ E o tempo médio de sobrevida?

√ Comando R:

```
> survfit( formula = Surv(tempo, status)~1, data = ipec90)
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec90)

records   n.max n.start  events  median 0.95LCL 0.95UCL
     21      21      21      15      52      37      NA
```

52 é o tempo mínimo no qual se espera que praticamente 50% dos pacientes estejam vivos

√ Sumário do comando **`survfit`**:

```
time n.risk n.event survival s
  16     21       1   0.9524
  18     20       1   0.9048
  21     19       1   0.8571
  22     17       1   0.8067
  29     15       1   0.7529
  35     14       1   0.6992
  37     13       1   0.6454
  39     12       1   0.5916
  40     11       1   0.5378
  52      9       1   0.4781
  54      8       1   0.4183
  60      7       1   0.3585
```

$52 = \min\{t_j \mid S(t_j) \leq 0{,}50$

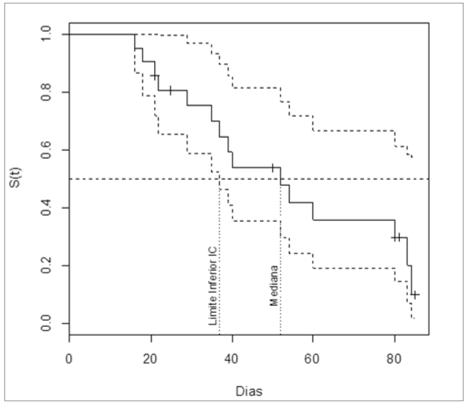
Após 52 dias, 8 pacientes estão vivos

√ Não equivalem a 50% dos pacientes

√ Espera-se que aconteça caso não houvesse censuras

- Intervalo de Confiança para t_{med}:
  - √ Há várias expressões para construir o intervalo de confiança
- Construção gráfica:
  - √ Tempo mediano corresponde à intersecção entre a linha S(t) = 0,50 e a curva da sobrevivência
  - √ Intervalo de confiança é dado pela intersecção da linha horizontal S(t) = 0,50 e as curvas do intervalos de confiança para S(t)

√ Tempo de sobrevivência mediano e intervalo de confiança de 95% dos pacientes com Aids



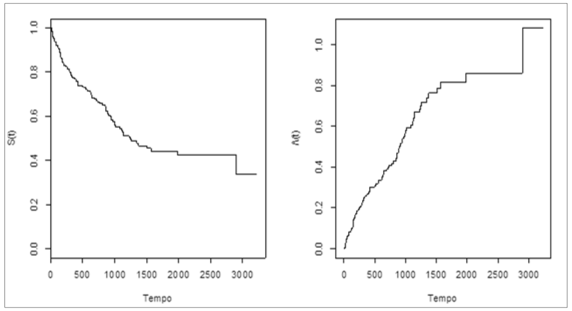
Estratificação

- Peculiaridades dos indivíduos contribuem para diferentes curvas de sobrevivências dos grupos
  - √ Pessoas não adoecem ou morrem da mesma forma
  - √ Descrição da sobrevivência segundo característica:
    - – Sexo, faixa etária, etc.
  - √ Divide-se o conjunto total de observações em estratos de acordo com covariável de interesse
  - √ Utilização de Kaplan-Meier para estimação da função de sobrevivência para cada estrato

Análise de Sobrevivência Aplicada à Saúde -- 2014

Exemplo

√ Coorte completa de Aids:
- – Óbitos: 90 e censuras: 103



√ Esta é uma boa estimativa da curva de sobrevivência tanto para homens quanto para mulheres?

Análise de Sobrevivência Aplicada à Saúde -- 2014

- Comandos no R:
  - √ Carregamento dados:

```
> library(survival)
> ipec <- read.table("dados/ipec.csv", header = T, sep = ";")
```

  - √ Gráfico coorte completa:

```
> par(mfrow = c(1,2))
> plot(aids.km, conf.int = F, xlab = "Tempo", ylab = "S(t)", mark.time = F)
> plot(aids.km, conf.int = F, xlab = "Tempo", ylab =
expression(paste(Lambda,"(t)")),
+ mark.time = F, fun = "cumhaz")
> par(mfrow = c(1,1))
```

Análise de Sobrevivência Aplicada à Saúde -- 2014

- Estratificação por sexo:

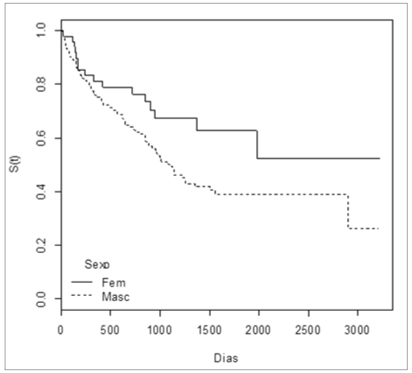
```
> survaids <- survfit(Surv(tempo, status)~sexo, data = ipec)
> survaids
Call: survfit(formula = Surv(tempo, status) ~ sexo, data = ipec)

       records n.max n.start events median 0.95LCL 0.95UCL
sexo=F      49    49      49     16     NA    1371      NA
sexo=M     144   144     144     74   1116     887    1563
```

  - √ Há predominância do sexo masculino (75%) e 82% das mortes se concentram neste grupo
  - √ Não foi possível estimar o tempo de sobrevivência mediano para mulheres (67% estão vivas)
  - √ Nesse caso é comum utilizar outro percentil:
    - – Percentil 70: 905 (F) e 516 (M)

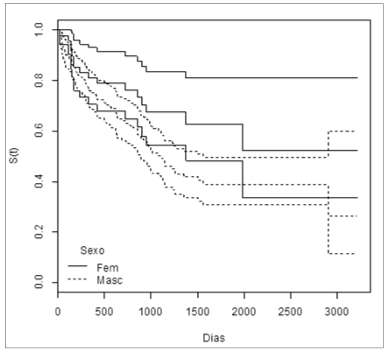
Análise de Sobrevivência Aplicada à Saúde -- 2014

- Curvas de sobrevivência por sexo



√ Curva de sobrevivência dos homens sempre abaixo
  - A diferença é significativa?

- Curvas de sobrevivência com IC de 95%



√ É visível a sobreposição das bandas de confiança
  - Sugere que a diferença observada não é significativa

Comparação das Curvas

- Testes de hipóteses:
  √ Log-Rank ou Mantel-Haenszel
  √ Peto
- Hipótese nula: não há diferença entre estratos
  √ H_0: $\lambda_1(t) = \lambda_2(t) = ... = \lambda_k(t)$

Teste *Log-rank*

- Distribuição esperada de eventos
  √ Sob H_0 é igual em todos os estratos:

$$E_k(t) = N(t)\frac{R_k(t)}{R(t)}$$

- Estatística de teste para 2 estratos (k = 2)

$$T_{LR} = \frac{(O_1 - E_1)^2}{\mathrm{Var}(O_1 - E_1)}$$

  √ O_1: total de eventos observados no estrato 1
  √ E_1: total de eventos esperados no estrato 1

- Fator de padronização da estatística (k = 2)

$$\mathrm{Var}(O_1 - E_1) = \sum_t \frac{R_1(t)R_2(t)\Delta N(t)[R(t) - \Delta N(t)]}{R^2(t)[R(t) - 1]}$$

- Distribuição amostral da estatística (caso geral)
  - √ Sob H_0 $\quad T_{LK} \overset{as.}{\sim} \chi^2_{k-1}$

Teste de Peto

- Maior peso às diferenças (ou semelhanças) no início da curva
  - √ Local de concentração da maior parte dos dados
- Estatística de teste: $T_P = \dfrac{(O_1 - E_1)^2}{\mathrm{Var}(O_1 - E_1)}$

  com $\quad \mathrm{Var}(O_1 - E_1) = \sum_{t_j} S(t_j)(O_1(t_j - E_1(t_j))$

  - √ Usa um ponderador S(t) no estimador
- Distribuição amostral
  - √ Sob H_0: $\quad T_P \overset{as.}{\sim} \chi^2_{k-1}$

Exemplo

- Coorte total do estudo sobre Aids
  - √ Teste Log-rank – Comandos e Saída no R:

```
> survdiff(Surv(tempo, status)~sexo, data = ipec, rho = 0)# Teste log-rank
Call:
survdiff(formula = Surv(tempo, status) ~ sexo, data = ipec)

         N Observed Expected (O-E)^2/E (O-E)^2/V
sexo=F  49       16     24.5      2.93      4.03
sexo=M 144       74     65.5      1.09      4.03

 Chisq= 4  on 1 degrees of freedom, p= 0.0447
```

  - √ Use rho = 0 (default)

- √ Teste Peto – Comandos e Saída no R:

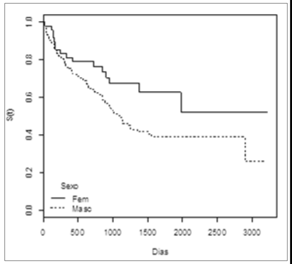
```
> survdiff(Surv(tempo, status)~sexo, data = ipec, rho = 1)# Teste Peto
Call:
survdiff(formula = Surv(tempo, status) ~ sexo, data = ipec, rho = 1)

         N Observed Expected (O-E)^2/E (O-E)^2/V
sexo=F  49     12.1     18.2     2.011      3.54
sexo=M 144     55.1     49.0     0.746      3.54

 Chisq= 3.5  on 1 degrees of freedom, p= 0.0598
```

- √ Use rho = 1

Comentários



- Resultados

Teste	Estatística	p-valor
Log-rank	4,03	0,0447
Peto	3,54	0,0598

√ Teste Log-rank rejeita H_0
(ainda que a diferença seja pequena)

√ Inspeção gráfica:
- Diferença nas curvas é maior mais tardiamente
- Teste Peto não detecta essa diferença (mais peso para os tempos menores)

Exemplo

- Estudo experimental com camundongos para verificação da imunização pela malária

√ 44 camundongos infectados pela malária foram alocados aleatoriamente em 3 grupos
- Grupo 1: imunizados 30 dias antes da infecção e infectados também pela esquistossomose
- Grupo 2: sem imunização
- Grupo 3: sem imunização e infectados também pela esquistossomose

√ Resposta de interesse:
- Tempo decorrido até a morte (dias)

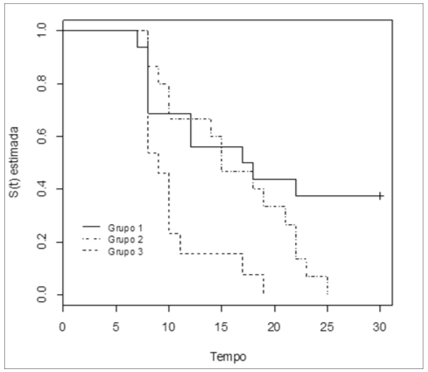
√ Duração do estudo: 30 dias

- Tempo observados no estudo da 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											

√ Exemplo caracterizado por censura do tipo I

- Sobrevivência estimada por Kaplan-Meier



√ Há diferenças entres os grupos?

- Teste log-rank para os três grupos

```
>  survdiff(Surv(tempos,cens)~grupo,rho=0)# Diferença todos os grupos
Call:
survdiff(formula = Surv(tempos, cens) ~ grupo, rho = 0)

        N Observed Expected (O-E)^2/E (O-E)^2/V
grupo=1 16       10    17.00    2.8816    6.4111
grupo=2 15       15    14.51    0.0167    0.0317
grupo=3 13       13     6.49    6.5190   10.4447

 Chisq= 12.6  on 2 degrees of freedom, p= 0.00187
```

√ Peto:
  - χ^2= 7,7 e p-valor = 0,0215

√ Indica diferença entre os grupos

√ Quais curvas diferem entre si?

Comparações Múltiplas

√ Em planejamento de experimentos, com hipótese de modelo linear com resposta normal:
  - Há vários métodos disponíveis para a realização de tais comparações

√ O mesmo não acontece com dados de sobrevivência

√ Possibilidade:
  - Comparar os grupos dois a dois, controlando o erro tipo I pelo método Bonferroni

Comparações Múltiplas

- Se H_0 é rejeitada, deseja-se saber quais são as diferenças

√ Procedimentos para fazer essas comparações são denominados Métodos de Comparações Múltiplas

√ Sejam:
  - α_{total}: probabilidade de rejeitar incorretamente hipótese de igualdade entre sobrevivências em favor da hipótese de que pelo menos duas das curvas são diferentes.
  - $\alpha_{comparação}$: probabilidade de rejeitar incorretamente comparação individual entre duas curvas de sobrevivência

- Considerando duas comparações
  - E_i: {Rejeição igualdade *i*}
  - Se $P(E_1|H_0\ V) = P(E_2|\ H_0\ V) = \alpha$

√ Pelo menos uma rejeição de igualdade, H_0 verdadeira

$$\mathrm{P}(E_1 \cup E_2|H_0\,V) = \mathrm{P}(E_1|H_0\,V) + \mathrm{P}(E_2|H_0\,V) - \mathrm{P}(E_1 \cap E_2|H_0\,V)$$
$$= \alpha + \alpha - \alpha^2$$

ou $$\mathrm{P}(E_1 \cup E_2|H_0 V) = 1 - \mathrm{P}(E_1^c \cap E_2^c|H_0 V)$$
$$= 1 - (1-\alpha)^2$$

- Se são executadas *n* comparações independentes

$$\alpha_{\text{total}} = 1 - (1 - \alpha_{\text{comparação}})^n$$

- Se não assumirmos que as comparações são independentes: $\alpha_{\text{total}} \leq n\, \alpha_{comparação}$

Correção de Bonferroni

- Para assegurar que o nível de significância de todas as comparações seja pelo menos α_{total}
  - √ Ajusta-se $\alpha_{\text{comparação}} = \dfrac{\alpha_{\text{total}}}{n}$
  - √ É altamente conservativo
    - Assegura que o máximo nível de significância global seja α
  - √ Produz estimativas menos precisas, mas limita a probabilidade de rejeições incorretas a um máximo de α_{total}.
  - √ É usado com pequeno número de comparações

- No nosso exemplo:
  - √ O nível de significância associado às 3 comparações é (3)(0,05) = 15%
  - √ Adota-se, 0,05/3 = 0,017 para cada um dos três testes
  - √ Pretende-se garantir uma conclusão geral a um nível de significância de no máximo 5%


```
>  survdiff(Surv(tempos[1:31],cens[1:31])~grupo[1:31],rho=0)
Call:
survdiff(formula = Surv(tempos[1:31], cens[1:31]) ~ grupo[1:31],
    rho = 0)

               N Observed Expected (O-E)^2/E (O-E)^2/V
grupo[1:31]=1 16       10     13.7      1.01      2.53
grupo[1:31]=2 15       15     11.3      1.23      2.53

 Chisq= 2.5  on 1 degrees of freedom, p= 0.112
```

Grupos 1 x 2

```
>  survdiff(Surv(tempos[17:44],cens[17:44])~grupo[17:44],rho=0)
Call:
survdiff(formula = Surv(tempos[17:44], cens[17:44]) ~ grupo[17:44],
    rho = 0)

                N Observed Expected (O-E)^2/E (O-E)^2/V
grupo[17:44]=2 15       15    20.53      1.49      7.98
grupo[17:44]=3 13       13     7.47      4.08      7.98

 Chisq= 8  on 1 degrees of freedom, p= 0.00473
```

Grupos 2 x 3

```
>  survdiff(Surv(c(tempos[1:16],tempos[32:44]),c(cens[1:16],
+          cens[32:44]))~c(grupo[1:16],grupo[32:44]),rho=0)
Call:
survdiff(formula = Surv(c(tempos[1:16], tempos[32:44]), c(cens[1:16],
    cens[32:44])) ~ c(grupo[1:16], grupo[32:44]), rho = 0)

                                N Observed Expected (O-E)^2/E (O-E)^2/V
c(grupo[1:16], grupo[32:44])=1 16       10    15.34      1.86      7.86
c(grupo[1:16], grupo[32:44])=3 13       13     7.66      3.72      7.86

 Chisq= 7.9  on 1 degrees of freedom, p= 0.00504
```

Grupos 1 x 3

- Teste log-rank

Grupos comparados	Estatística de teste	p-valor
1x2	2,5	0,112
2x3	8,0	0,005
1x3	7,9	0.009

  - √ Diferença significativa entre os grupos 1x3 e 2x3
  - √ Não foram encontradas evidências de diferenças entre os grupos 1x2

Conclusões

- Diferença 1 e 3:
  - √ Atesta eficácia de imunização pela malária na presença de infecções pela malária e pela esquistossomose
- Diferença 2 e 3:
  - √ Mostra impacto na mortalidade de camundongs devido à infecção pela esquistossomose

Referências

Bibliografia

- Carvalho, M. S. et al. *Análise de Sobrevivência: Teoria e Aplicações em Saúde*. (Fiocruz)
- Colosimo, E. A. e Giolo, S. R. *Análise de Sobrevivência Aplicada*. (Edgard Blucher)
- Klein, J. P. e Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. (Springer)
- Kleinbaum, D. G. *Survival Analysis: a Self-Learning Text*

Material de Apoio

- R:
  - √ www.r-project.org
- Tutorial online do R:
  - √ http://www.leg.ufpr.br/Rtutorial
  - √ http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa
- Conjuntos de dados e material Análise de Sobrevivência – Carvalho et al.
  - √ http://sobrevida.fiocruz.br