Outline

Estimação Não-Paramétrica

- Estimadores de sobrevida e risco
- Kaplan-Meier e Nelson Aalen
- Intervalos de confiança
- Kaplan-Meier estratificado
- Testes de Log-Rank e Peto

Incorporando a censura

Sem suposições sobre a distribuição do tempo

Estimação Não-Paramétrica

- Estimadores de sobrevida e risco
- Kaplan-Meier e Nelson Aalen
- Intervalos de confiança
- Kaplan-Meier estratificado
- Testes de Log-Rank e Peto

Incorporando a censura

Sem suposições sobre a distribuição do tempo

Kaplan-Meier

- A probabilidade de sobrevida até o tempo t é estimada considerando que a sobrevivência até cada tempo é independente da sobrevivência até outros tempos.
- A probabilidade de chegar até o tempo t é o produto da probabilidade de chegar até cada um dos tempos anteriores.

Kaplan-Meier

- Seja $t_1 < t_2 < \cdots < t_m$ os tempos onde ocorreram os eventos;
- $Y_i(t) = 1$ se a pessoa i está em risco no tempo t e 0 caso contrário.
- $R(t_i)$ é o total de pessoas a risco no tempo t_i.
- A cada tempo t_i em que houver um evento, a probabilidade de sobrevivência será o número dos que sobreviveram até aquele tempo $(R(t_i) - N(t_i))$ sobre os que estavam em risco naquele tempo $(R(t_i))$.
- O estimador da distribuição $S(t)$ é o produto das probabilidades de sobrevivência a cada tempo $t_i \leq t$.

Kaplan-Meier

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

ou na forma de produtório:

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \frac{R(t_i) - N(t_i)}{R(t_i)}$$

Da sobrevida ao risco

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

Logo.... pode-se estimar qualquer das funções.

Estimador de Nelson-Aalen

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

Melhor para amostras muito pequenas

planilha exerciciokm.ods

Intervalos de confiança

Variância do estimador Kaplan-Meier para a sobrevida
Estimador de Greenwood

$$Var(\hat{S}_{KM}(t)) = (\hat{S}_{KM}(t))^2 \sum_{t_i \leq t} \frac{N(t_i)}{R(t_i)(R(t_i) - N(t_i))}$$

Intervalos de confiança

Assumindo erro α, o intervalo fica assim:

$$\left[\hat{S}_{KM}(t) - z_{\alpha/2}\sqrt{Var(\hat{S}_{KM}(t))}; \hat{S}_{KM}(t) + z_{\alpha/2}\sqrt{Var(\hat{S}_{KM}(t))}\right]$$

Entretanto, este intervalo permite valores negativos e maiores do que 1, o que é incompatível com a definição de sobrevida.

Intervalos de confiança

Construindo intervalo simétrico para o risco $\ln \Lambda(t) = \ln(-\ln S(t))$, pode-se obter um intervalo assimétrico para $S(t)$, porém sempre positivo e menor do que 1 e igual a

$$[\exp(-\exp(l_s)); \exp(-\exp(l_i))]$$

onde

$$[l_i; l_s] = \left[\ln(\hat{\Lambda}_{KM}(t)) - z_{\alpha/2} dp; \ln(\hat{\Lambda}_{KM}(t)) + z_{\alpha/2} dp\right]$$

e o desvio padrão dp é:

$$dp = \sqrt{\frac{\sum_{t_i \leq t} \frac{N(t_i)}{R(t_i)(R(t_i) - N(t_i))}}{\left\{\sum_{t_i \leq t} \ln\left[\frac{R(t_i) - N(t_i)}{N(t_i)}\right]\right\}^2}}$$

no R

- Criando o objeto sobrevida (tempo, censura):

```
 > Surv(tempo,status)
# variável status=1 indica evento, 0 censura
```

```
16 18 21+ 21 22 25+ 29 35 37 39 40 50+ 52 54 60 80+ 80 81+ 83 84 85+
```

- Kaplan-Meier

```
> KM <- survfit(Surv(tempo,status), data = ipec90)
> summary(KM)
> plot(KM)
```

- Nelson-Aalen

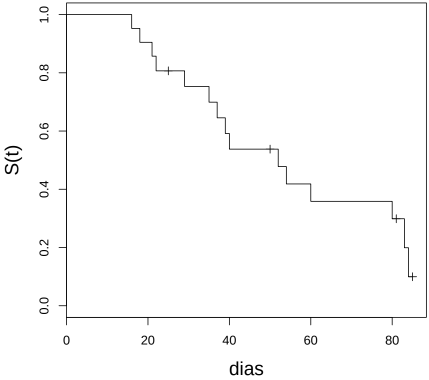
```
> sob.NA <- survfit(coxph(y~1, data = ipec90))
> sob.NA
> summary(sob.NA)
```

Saídas do R – summary(KM)

time	n.risk	n.event	survival	std.err	lower95%CI	upper95%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

Saídas do R – plot(KM)

Função de sobrevida dos pacientes com aids, utilizando o estimador produto Kaplan-Meier.
Os símbolos + localizam as censuras.



Kaplan-Meier estratificado

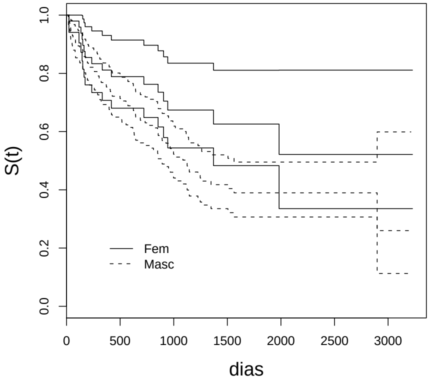
- A sobrevivência é estimada separadamente para cada estrato, utilizando Kaplan-Meier.
- no R

```
> ipec <- read.table("ipec.csv",header=T,sep=";")
> survaids <- survfit(Surv(tempo,status)~ sexo, data = ipec)
> survaids

Call: survfit(formula = resp ~ sexo, data = ipec)

         n events rmean se(rmean) median 0.95LCL 0.95UCL
sexo=F  49     16  2096       229    Inf    1371     Inf
sexo=M 144     74  1581       122   1116     887    1563
```

Gráfico sobrevida estratificada



Curvas de sobrevida de pacientes com aids, estratificado por sexo. Estimação por Kaplan-Meier, com intervalo de confiança de 95%.

Testes

Hipótese nula: não há diferença entre estratos

$$H_0 : \lambda_1(t) = \lambda_2(t) = \cdots = \lambda_k(t)$$

Log-rank (ou Mantel-Haenszel)

Distribuição esperada de eventos igual em todos os estratos:

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

Estatística de teste log-rank para dois estratos $(k = 2)$:

$$\text{Log-rank} = \frac{(N_1 - E_1)^2}{Var(N_1 - E_1)}$$

com $N_1 =$ ao total de eventos observados no estrato 1 e $E_1 =$ ao total de eventos esperados no estrato 1.

Teste log-rank

A variância, que entra no cálculo como um fator de padronização, tem a fórmula (para $k = 2$):

$$Var(N_1 - E_1) = v_i$$

em que

$$v_i = \sum_{t_i} \frac{R_1(t_i)[R(t_i) - R_1(t_i)]N(t_i)[R(t_i) - N(t_i)]}{R(t_i)^2[R(t_i) - 1]}.$$

A estatística log-rank, sob a hipótese nula, segue uma distribuição χ^2 , com $k - 1$ graus de liberdade.

Teste de Peto

Dá maior peso às diferenças (ou semelhanças), no início da curva, onde se concentra a maior parte dos dados e por isso é mais informativa. Usa um ponderador $S(t)$ no estimador.

$$\text{Peto} = \frac{(N_1 - E_1)^2}{Var(N_1 - E_1)}$$

sendo que

$$N_1 - E_1 = \frac{\sum S(t_i)(N_1(t_i) - E_1(t_i))}{\sum S(t_i)}$$

$$Var(N_1 - E_1) = \frac{(\sum S(t_i)(N_1(t_i) - E_1(t_i)))^2}{\sum (S(t_i))^2 v_i}$$

Também a estatística Peto segue aproximadamente uma distribuição χ^2 com $k - 1$ graus de liberdade.

no R

```
> survdiff(Surv(tempo,status)~sexo, data=ipec,rho=0)

Call:
survdiff(formula = Surv(tempo, status) ~ sexo, data = ipec, rho = 0)

         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
```

O argumento *rho* determina o tipo de teste a ser realizado. Para log-rank, use *rho = 0* (*default*). Para o teste Peto, use *rho = 1*.

no R

```
> survdiff(Surv(tempo,status)~sexo, data=ipec,rho=1)

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
```