EPI10 - Análise de Sobrevivência

Modelo de Cox

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Faculdade de Medicina
Programa de Pós-Graduação em Epidemiologia

Porto Alegre, 2022

Estimação: o método da máxima verossimilhança (em poucas palavras)

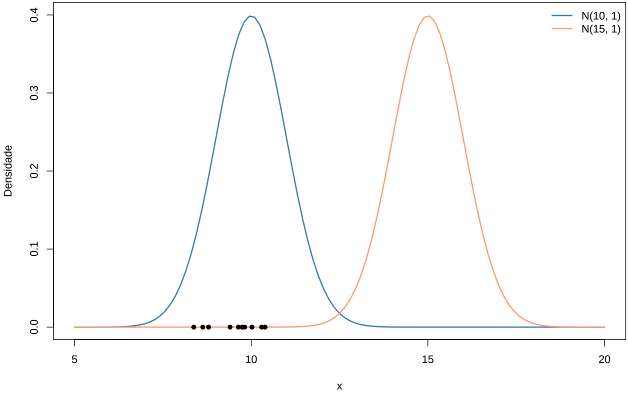
Método da máxima verossimilhança

O **método da máxima verossimilhança** trata o problema de estimação da seguinte forma:

- Com base nos resultados observados em uma amostra, $x_1, x_2, \ldots, x_n$, qual é a distribuição (modelo probabilístico), entre todas aquelas definidas pelos possíveis valores de seus parâmetros, com maior **verossimilhança** de ter gerado a amostra?

Método da máxima verossimilhança

Exemplo. Qual dos dois modelos abaixo tem maior verossimilhança de ter gerado a amostra?



Método da máxima verossimilhança

- Formalizando, dada uma amostra de observações independentes, $x_1, x_2, \ldots, x_n$, e com mesma distribuição, parametrizada pelo parâmetro genérico θ, $f(x_i; \theta)$, a **função de verossimilhança** é definida como:

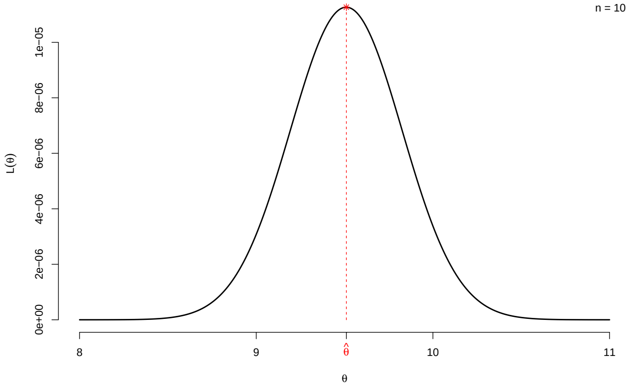
$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

- A **estimativa de máxima verossimilhança (EMV)** do parâmetro θ, $\widehat{\theta}$ é o valor de θ que maximiza $L(\theta)$. Ou seja,

$$\widehat{\theta} = \arg\max_{\theta} L(\theta).$$

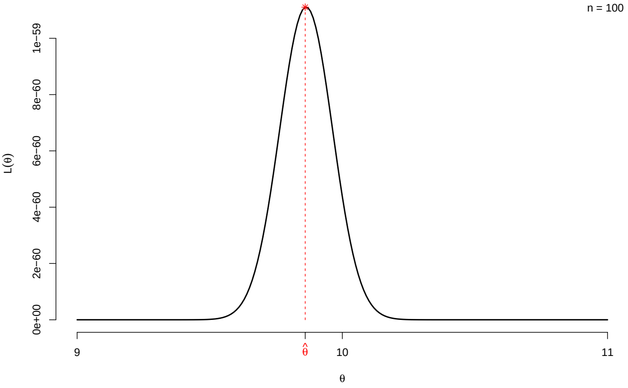
Método da máxima verossimilhança

Visualmente, temos



Método da máxima verossimilhança

Quando *n* aumenta



Método da máxima verossimilhança

Comentários

- Avaliar o ponto que maximiza $L(\theta)$ é equivalente a avaliar o ponto que maximiza $\ell(\theta) = \log L(\theta)$.
- θ pode ser um vetor de parâmetros.
- Em geral, a obtenção do ponto de máximo da função de verossimilhança se dá por meio de **métodos numéricos**.

Método da máxima verossimilhança

Principais propriedades dos estimadores de máxima verossimilhança

- Distribuição assintoticamente normal.
  - Ou seja, para grandes amostras, $\widehat{\theta} \overset{a}{\sim} N(\theta, \sigma^2_{\widehat{\theta}})$.
- Propriedade da invariância.
  - Se $\widehat{\theta}$ é EMV para θ, então $e^{\hat{\theta}}$ é EMV para e^{θ}, por exemplo.
- A partir da normalidade assintótica, podemos
  - Construir intervalos de confiança;
  - Realizar testes de hipoóteses (Wald, Score, Razão de verossimilhanças).

Função de verossimilhança em dados de sobrevivência

- Dada uma amostra de observações independentes, $(t_1, \delta_1), (t_2, \delta_2), \ldots, (t_n, \delta_n)$, e assumindo que o **mecanismo de censura é não-informativo**[1], a função de verossimilhança é dada por

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} [f(t_i;\theta)]^{\delta_i} \times [S(t_i;\theta)]^{1-\delta_i} \\
&= \prod_{i=1}^{n} [\lambda(t_i;\theta)]^{\delta_i} \, S(t_i;\theta).
\end{aligned}
$$

[1] A censura não é informativa, no caso em que, quando esta ocorre não dá nenhuma informação sobre o evento em estudo e as informações provenientes dos pacientes perdidos no seguimento são ignoradas. Essa hipótese de a censura ser independente do evento é uma hipótese forte que raramente é verificada. Por exemplo, no HIV, os pacientes que perdem o acompanhamento são geralmente os casos mais graves. A violação desta suposição acarreta em vieses.

Modelo de Cox

Modelos de regressão em sobrevivência

Duas classes de modelos de regressão se destacam em análise de sobrevivência:

- Modelos de **tempos de vida acelerados** ou **modelos paramétricos**.
- Modelos de **riscos (taxas de falha) proporcionais** ou **modelo semiparamétrico de Cox**.

Modelo de regressão de Cox

O modelo de riscos proporcionais, também chamado de modelo de Cox (Cox, 1972)[2]

- abriu uma nova fase na modelagem de dados clínicos;
- é o mais utilizado na análise de dados de sobrevivência;
- permite incorporar facilmente covariáveis dependentes do tempo, que ocorrem com frequência em estudos clínicos e epidemiológicos.

---

[2] Cox, D.R. (1972), Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34: 187-202.

Modelo de regressão de Cox

- Assume-se, nesse modelo, que os tempos t_i , $i = 1, \ldots, n$, são independentes e que a **taxa de falha (risco)** tem a seguinte forma:

$$\lambda(t) = \lambda_0(t)\exp\{\beta_1 x_1 + \ldots + \beta_p x_p\}.$$

- O componente não-paramétrico, $\lambda_0(t)$, **não é especificado** e é uma função não-negativa do tempo.
  - Ele é usualmente chamado de **função de taxa de falha basal**.
- O componente paramétrico $\exp\{x'\beta\} = \exp\{\beta_1 x_1 + \ldots + \beta_p x_p\}$ é o nosso interesse, em especial no vetor de parâmetros β, e $x' = (x_1, x_2, \ldots, x_p)$ **é um vetor de covariáveis observadas** (como, por exemplo, **sexo**, **idade**, **grupo de tratamento ou espoxição**, etc.).

Riscos proporcionais

- O modelo é conhecido por ter taxas de falha (riscos) proporcionais.
  - Este fato é conveniente na sua interpretação.
- Ou seja, a razão das taxas de falha de dois indivíduos diferentes, i e j, é

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp\{x_i{}'\beta\}}{\lambda_0(t)\exp\{x_j{}'\beta\}} = \exp\{x_i{}'\beta - x_j{}'\beta\},$$

que **não depende do tempo t**.

Riscos proporcionais

- Assim, se um indivíduo no início do estudo tem um risco de óbito igual a duas vezes o risco de um segundo indivíduo, então esta razão de riscos será a mesma para todo o período de acompanhamento.
- Note que o modelo assume que as funções de taxa de falha de indivíduos distintos difere apenas com respeito às covariáveis destes indivíduos.

Riscos proporcionais

- Ainda considerando a propriedade de riscos proporcionais, suponha um estudo controlado que consiste na comparação dos tempos de falha de dois grupos em que os pacientes são selecionados aleatoriamente para receber o tratamento padrão (grupo controle, $x = 0$) ou o novo tratamento ($x = 1$), tem-se:

$$\frac{\lambda_1(t)}{\lambda_0(t)} = K,$$

em que K é uma razão de taxas de falha, constante para todo tempo t de acompanhamento do estudo.

Riscos proporcionais

- Se $K = \exp\{\beta x\}$, temos o modelo de Cox para um única covariável dicotômica:

$$\lambda(t) = \lambda_0(t) \exp\{\beta x\},$$

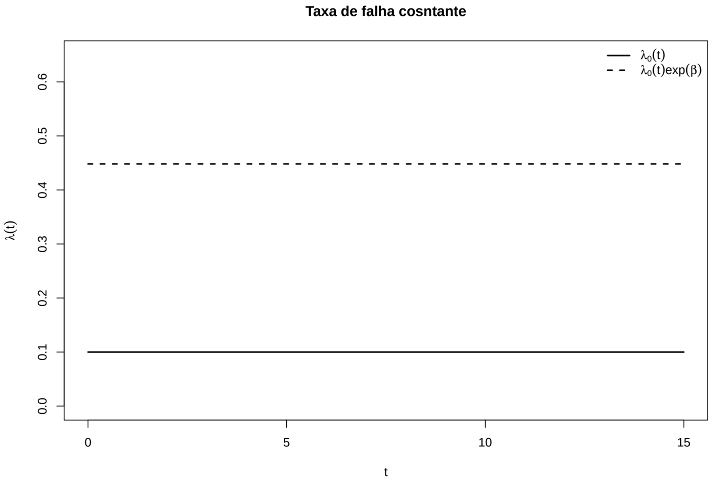
ou seja,

$$\lambda(t) = \begin{cases} \lambda_1(t) = \lambda_0(t) \exp\{\beta x\}, & \text{se} \quad x = 1, \\ \lambda_0(t), & \text{se} \quad x = 0. \end{cases}$$

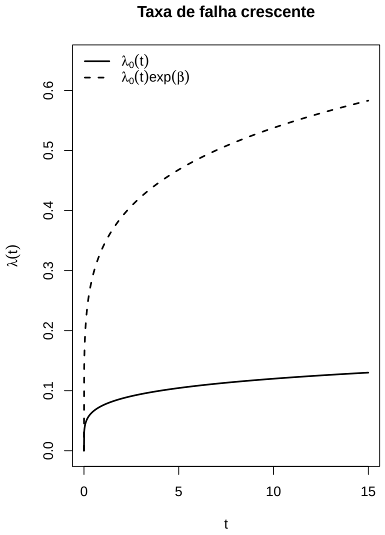
Riscos proporcionais

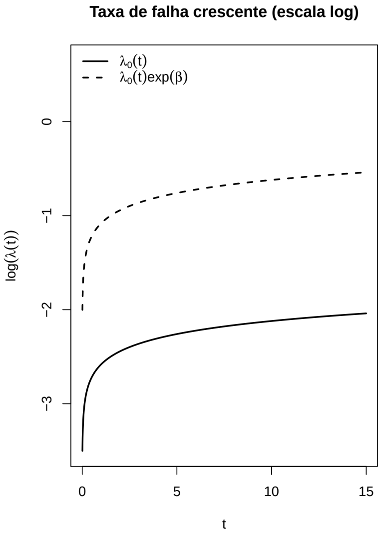
Suponha $\beta > 0$. Graficamente, temos que o novo tratamento **acelera** a taxa de falha dos pacientes.

Riscos proporcionais

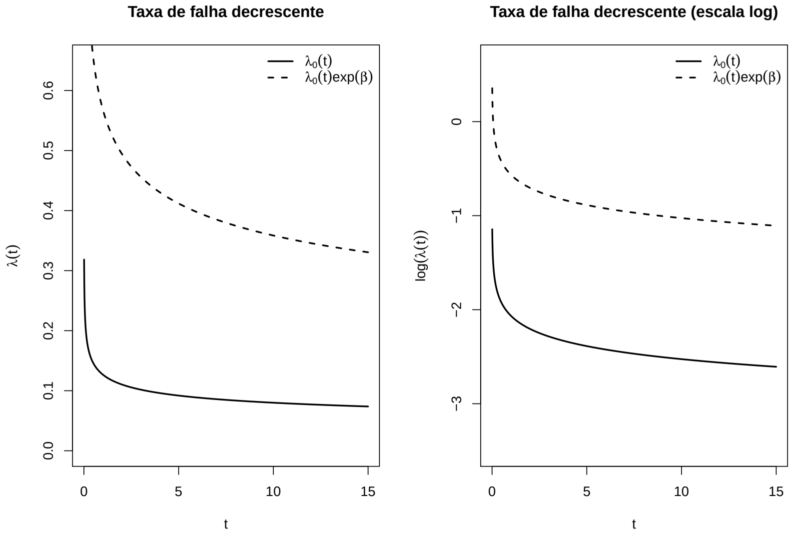


Riscos proporcionais





Riscos proporcionais

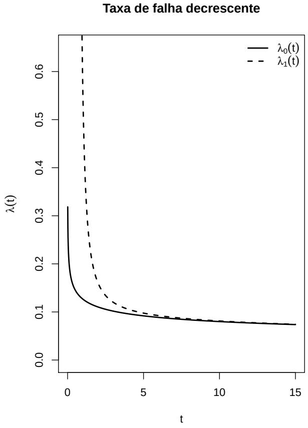


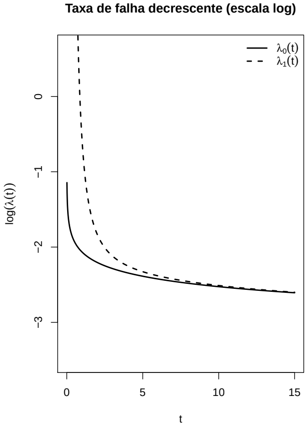
Riscos proporcionais

- Desta série de exemplos da proporcionalidade dos riscos do modelo de Cox, reforça-se que o foco deste modelo não está na forma que $\lambda_0(t)$ assume, mas sim no **efeito (multiplicativo) que a(s) covariável(eis)** tem em $\lambda_0(t)$.

Riscos proporcionais

- Por fim, apresentamos um exemplo gráfico em que a proporcionalidade dos riscos **não se verifica**.





Estimação no modelo de Cox

- O modelo de regressão de Cox é caracterizado pelos coeficientes $\beta' s$, que **medem os efeitos das covariáveis sobre a função de taxa de falha**.
- Um método de estimação é necessário para se fazer inferências no modelo.
- O método de máxima verossimilhança não é adequado devido a presença do componente não-paramétrico $\lambda_0(t)$.

$$L(\beta) = \prod_{i=1}^{n} [\lambda_0(t_i) \exp\{x_i' \beta\}]^{\delta_i} [S_0(t_i)]^{\exp\{x_i' \beta\}}.$$

Função de verossimilhança parcial

- Cox (1975)[3] propôs então uma solução alternativa: **verossimilhança parcial**.
- Este método consiste em condicionar a construção da função de verossimilhança ao conhecimento da história passada de falhas e censuras.
- Desta forma, elimina-se o componente não-paramétrico da função de verossimilhança.

[3] Cox, D. R., Partial likelihood, *Biometrika*, Volume 62, Issue 2, August 1975, Pages 269–276.

Função de verossimilhança parcial

- A **função de verossimilhança parcial** é utilizada para fazer inferência no modelo de Cox.
- A verossimilhança parcial é formada pelo produto de todos os indivíduos da amostra:

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp\{x_i'\beta\}}{\sum_{j \in R(t_i)} \exp\{x_j'\beta\}} = \prod_{i=1}^{n} \left(\frac{\exp\{x_i'\beta\}}{\sum_{j \in R(t_i)} \exp\{x_j'\beta\}} \right)^{\delta_i},$$

em que $R(t_i)$ **é o conjunto dos índices das observações sob risco no tempo** t_i, e δ_i é indicadora de falha ou censura.

- Os valores de β **que maximizam** $L(\beta)$ **são as estimativas de máxima verossimilhança parcial**, $\widehat{\beta}$, de β.

Função de verossimilhança parcial

- A função de verossimilhança parcial proposta por Cox assume que os tempos de ocorrência de evento são distintos (ausência de tempos empatados).
- Para o caso de tempos de falha empatados, uma aproximação da função de verossimilhança deve ser usada.
- Entre as diversas propostas na literatura, as aproximações propostas por **Breslow (1972)**[4] e **Efron (1977)**[5] são as que mais comumente são encontradas nas implementações dos *softwares* de análise de sobrevivência.

---

[4] Breslow, N. (1972), Discussion on Professor Cox's Paper. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34: 202-220.

[5] Bradley Efron (1977) The Efficiency of Cox's Likelihood Function for Censored Data, *Journal of the American Statistical Association*, 72:359, 557-565.

Inferência no modelo de Cox

- Assim como os EMV, os estimadores de máxima verossimilhança parcial, em amostras grandes, tem distribuição aproximadamente normal.
- Isto nos permite construir **intervalos de confiança** $100 \times (1-\alpha)\%$ para β_r **(o efeito da r-ésima covariável)**

$$\hat{\beta}_r \pm z_{1-\alpha/2}\widehat{ep}(\widehat{\beta}_r),$$

em que $z_{1-\alpha/2}$ é quantil correspondente na distribuição normal padrão, e $\widehat{ep}(\widehat{\beta}_r)$ é a estimativa do erro padrão de $\widehat{\beta}_r$.

Inferência no modelo de Cox

- Também podemos construir um teste de hipóteste para avaliar $H_0 : \beta_r = 0$. O **teste de Wald** é bastante utilizado

$$z = \frac{\widehat{\beta}_r}{\widehat{ep}(\widehat{\beta}_r)} \overset{a}{\sim} N(0,1).$$

- O **teste da razão de verossimilhanças** geralmente é usado para comparação de modelos.

Interpretação dos coeficientes estimados

- O efeito das covariáveis é de acelerar ou desacelerar a função de risco.
- A propriedade de taxas proporcionais é extremamente útil na interpretação dos coeficientes estimados.
- A razão das taxas de falha de dois indivíduos i e j que têm os mesmos valores para as covariáveis com exceção da l-ésima, tem-se

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \exp\{\beta_l(x_{il} - x_{jl})\},$$

que é interpretado como a **razão de taxas de falha (*hazard ratio*, *HR*; ou ainda "razão de riscos")**.

Interpretação dos coeficientes estimados

- Por exemplo, suponha que x_l seja uma covariável dicotômica indicando pacientes hipertensos.
  - A taxa de óbito entre os hipertensos é $\exp(\beta_l)$ **vezes** a taxa daqueles com pressão normal, **mantida fixas as outras covariáveis**.
- Uma interpretação similar é obtida para covariáveis contínuas.
  - Se, por exemplo, o efeito de idade (em anos) é $e^{\beta} = 1,05$ para este termo, tem-se com o **aumento de 1 ano na idade, que a taxa de óbito aumenta em 5%**.

Interpretação dos coeficientes estimados

- Estimativa pontual para $HR = \exp\{\beta_l\}$ pode ser obtida utilizando-se a **propriedade de invariância** do estimador de máxima verossimilhança parcial.
- Para obtenção da **estimativa intervalar**, é necessário obter uma **estimativa do erro padrão** de $\exp\{\widehat{\beta}_l\}$.
  - Isto pode ser feito utilizando-se o **método delta**.
- Retornando ao exemplo dos pacientes hipertensos e com pressão normal:
  - O valor 1 pertencendo ao intervalo estimado, indica não haver evidências de que os riscos dos pacientes hipertensos e com pressão normal apresentem diferenças significativas.

Exemplo

Estudo sobre câncer de laringe

Neste exemplo, os dados considerados referem-se a um estudo realizado com 90 pacientes do sexo masculino diagnosticados no período de 1970 a 1978 com câncer de laringe e que foram acompanhados até 01/01/1983.

Para cada paciente, foram registrados, no diagnóstico:

- a idade (em anos);
- o estágio da doença (ordenados por grau de severidade da doença): I. tumor primário;
  - **II.** envolvimento de nódulos;
  - **III.** metástases;
  - **IV.** combinações dos 3 estágios anteriores.
- tempos de óbito ou censura (em meses).

Estudo sobre câncer de laringe

```
df.laringe <- read.table(file = "../dados/laringe.txt",
                         header = TRUE)
```

```
head(df.laringe)
```

```
##   id tempos cens idade estagio
## 1  1    0.6    1    77       1
## 2  2    1.3    1    53       1
## 3  3    2.4    1    45       1
## 4  4    3.2    1    58       1
## 5  5    3.3    1    76       1
## 6  6    3.5    1    43       1
```

Estudo sobre câncer de laringe

```
str(df.laringe)
```

```
## 'data.frame':    90 obs. of  5 variables:
##  $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ tempos : num  0.6 1.3 2.4 3.2 3.3 3.5 3.5 4 4 4.3 ...
##  $ cens    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ idade   : int  77 53 45 58 76 43 60 52 63 86 ...
##  $ estagio: int  1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(df.laringe)
```

```
##        id            tempos            cens            idade          estagio
##  Min.   : 1.00   Min.   : 0.100   Min.   :0.0000   Min.   :41.00   Min.   :1.000
##  1st Qu.:23.25   1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.:57.00   1st Qu.:1.000
##  Median :45.50   Median : 4.000   Median :1.0000   Median :65.00   Median :2.000
##  Mean   :45.50   Mean   : 4.198   Mean   :0.5556   Mean   :64.61   Mean   :2.222
##  3rd Qu.:67.75   3rd Qu.: 6.200   3rd Qu.:1.0000   3rd Qu.:72.00   3rd Qu.:3.000
##  Max.   :90.00   Max.   :10.700   Max.   :1.0000   Max.   :86.00   Max.   :4.000
```

Estudo sobre câncer de laringe

```
df.laringe$estagio <- factor(x = df.laringe$estagio,
                             levels = 1:4,
                             labels = c("I", "II", "III", "IV"))

str(df.laringe)
```

```
## 'data.frame':    90 obs. of  5 variables:
##  $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ tempos  : num  0.6 1.3 2.4 3.2 3.3 3.5 3.5 4 4 4.3 ...
##  $ cens    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ idade   : int  77 53 45 58 76 43 60 52 63 86 ...
##  $ estagio: Factor w/ 4 levels "I","II","III",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(df.laringe)
```

```
##        id            tempos            cens            idade         estagio
##  Min.   : 1.00   Min.   : 0.100   Min.   :0.0000   Min.   :41.00   I  :33
##  1st Qu.:23.25   1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.:57.00   II :17
##  Median :45.50   Median : 4.000   Median :1.0000   Median :65.00   III:27
##  Mean   :45.50   Mean   : 4.198   Mean   :0.5556   Mean   :64.61   IV :13
##  3rd Qu.:67.75   3rd Qu.: 6.200   3rd Qu.:1.0000   3rd Qu.:72.00
##  Max.   :90.00   Max.   :10.700   Max.   :1.0000   Max.   :86.00
```

Estudo sobre câncer de laringe

```r
ekm <- survfit(Surv(time = tempos, event = cens) ~ estagio,
               data = df.laringe,
               conf.type = "log-log")

ekm
```

```
## Call: survfit(formula = Surv(time = tempos, event = cens) ~ estagio,
##     data = df.laringe, conf.type = "log-log")
##
##              n events median 0.95LCL 0.95UCL
## estagio=I   33     15    6.5     4.3      NA
## estagio=II  17      7    7.0     3.6      NA
## estagio=III 27     17    5.0     1.6     7.8
## estagio=IV  13     11    1.5     0.4     3.6
```

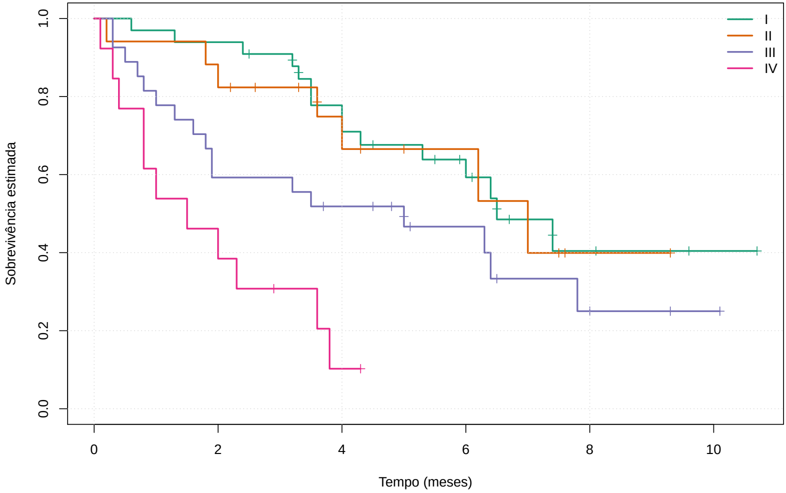
Estudo sobre câncer de laringe

```
plot(ekm, conf.int = FALSE,
     mark.time = TRUE,
     col = c("#1B9E77", "#D95F02",
             "#7570B3", "#E7298A"),
     lwd = 2, xlab = "Tempo (meses)",
     ylab = "Sobrevivência estimada")

abline(h = seq(0, 1, by = 0.2),
       v = seq(0, 16, by = 4),
       col = "lightgrey", lty = 3)

legend("topright",
       c("I", "II", "III", "IV"),
       col = c("#1B9E77", "#D95F02",
             "#7570B3", "#E7298A"),
       lwd = 2, bty = "n")
```

Estudo sobre câncer de laringe



Estudo sobre câncer de laringe

```
survdiff(Surv(time = tempos, event = cens) ~ estagio,
              data = df.laringe)
```

```
## Call:
## survdiff(formula = Surv(time = tempos, event = cens) ~ estagio,
##     data = df.laringe)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## estagio=I   33       15    22.57     2.537     4.741
## estagio=II  17        7    10.01     0.906     1.152
## estagio=III 27       17    14.08     0.603     0.856
## estagio=IV  13       11     3.34    17.590    19.827
##
##  Chisq= 22.8  on 3 degrees of freedom, p= 5e-05
```

Estudo sobre câncer de laringe

```
mod1 <- coxph(Surv(time = tempos, event = cens) ~ estagio,
              data = df.laringe, method = "breslow")

summary(mod1)
```

```
## Call:
## coxph(formula = Surv(time = tempos, event = cens) ~ estagio,
##     data = df.laringe, method = "breslow")
##
##   n= 90, number of events= 50
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## estagioII    0.06576   1.06797  0.45844 0.143   0.8859
## estagioIII   0.61206   1.84423  0.35520 1.723   0.0849 .
## estagioIV    1.72284   5.60040  0.41966 4.105 4.04e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## estagioII        1.068     0.9364    0.4348     2.623
## estagioIII       1.844     0.5422    0.9193     3.700
## estagioIV        5.600     0.1786    2.4604    12.748
```

Estudo sobre câncer de laringe

```
##
## Concordance= 0.668  (se = 0.037 )
## Likelihood ratio test= 16.26  on 3 df,   p=0.001
## Wald test            = 18.95  on 3 df,   p=3e-04
## Score (logrank) test = 22.46  on 3 df,   p=5e-05
```

Estudo sobre câncer de laringe

```
mod2 <- coxph(Surv(time = tempos, event = cens) ~ idade,
                 data = df.laringe, method = "breslow")

summary(mod2)
```

```
## Call:
## coxph(formula = Surv(time = tempos, event = cens) ~ idade, data = df.laringe
##     method = "breslow")
##
##   n= 90, number of events= 50
##
##          coef exp(coef) se(coef)     z Pr(>|z|)
## idade 0.02318   1.02345  0.01447 1.602    0.109
##
##       exp(coef) exp(-coef) lower .95 upper .95
## idade     1.023     0.9771    0.9948     1.053
##
## Concordance= 0.555  (se = 0.046 )
## Likelihood ratio test= 2.61  on 1 df,   p=0.1
## Wald test            = 2.57  on 1 df,   p=0.1
## Score (logrank) test = 2.58  on 1 df,   p=0.1
```

Estudo sobre câncer de laringe

```
mod3 <- coxph(Surv(time = tempos, event = cens) ~ estagio + idade,
              data = df.laringe, method = "breslow")

summary(mod3)
```

```
## Call:
## coxph(formula = Surv(time = tempos, event = cens) ~ estagio +
##     idade, data = df.laringe, method = "breslow")
##
##   n= 90, number of events= 50
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## estagioII  0.13856   1.14862  0.46231 0.300    0.764
## estagioIII 0.63835   1.89335  0.35608 1.793    0.073 .
## estagioIV  1.69306   5.43607  0.42221 4.010 6.07e-05 ***
## idade      0.01890   1.01908  0.01425 1.326    0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## estagioII      1.149     0.8706    0.4642     2.842
## estagioIII     1.893     0.5282    0.9422     3.805
```

Estudo sobre câncer de laringe

```
## estagioIV      5.436     0.1840    2.3763    12.436
## idade          1.019     0.9813    0.9910     1.048
##
## Concordance= 0.682  (se = 0.039 )
## Likelihood ratio test= 18.07  on 4 df,   p=0.001
## Wald test            = 20.82  on 4 df,   p=3e-04
## Score (logrank) test = 24.33  on 4 df,   p=7e-05
```

Para casa

1. Leia o capítulo 5 do livro **Análise de sobrevivência aplicada**[6].
2. Leia os capítulo 6 do livro **Análise de sobrevivência: teoria e aplicações em saúde**[7].

[6] Colosimo, E. A. e Giolo, S. R. **Análise de sobrevivência aplicada**, Blucher, 2006.
[7] Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S. e Shimakura, E. S. **Análise de sobrevivência: teoria e aplicações em saúde**, 2ª ed. Editora Fiocruz, 2011.

Próxima aula

- Estimação de funções relacionadas a $\lambda_0(t)$.
- Adequação do modelo de Cox.
  - Análise de resíduos.

Por hoje é só!

Bons estudos!