Introdução

Caio Gomes Alves

2023-03-29

Introdução

Seja y = tempo até o evento de interesse, chamamos esse tempo y de **tempo de falha**. A resposta para esses eventos é intrinsicamente longitudinal, sendo obtidas pot estudos observacionais e experimentais, de forma retrospectiva ou prospectiva.

Exemplos de estudos observacionais são:

- Estudos descritivos;
- Estudos caso-controle;
- Estudos de coorte.

Exemplos de estudos experimentais são principalmente:

- Estudos clínicos.

Num estudo de coorte, os grupos exposto e não-exposto ao fator de interesse são acompanhados no tempo, registrando-se a ocorrência do evento de interesse. Por exemplo, dois indivíduos são expostos ao cigarro (usuário x não-usuário), e são acompanhados até a ocorrência de câncer.

Denominações usuais para estudos de coorte são: * Estudos longitudinais ou de seguimento (ênfase no acompanhamento ao longo do tempo); * Estudos prospectivos (ênfase na direção do acompanhamento); * Estudos de incidência (atentando para a proporção de novos casos no período de seguimento).

Em relação à forma de coleta das informações, podemos classificá-las em:

- Coorte *contemporânea* ou *prospectiva*: indivíduos escolhidos no presente e desfecho registrado em acompanhamento futuro;

Um estudo *clínico aleatorizado* é considerado o padrão-ouro para testar a eficácia de uma intervenção. É um estudo prospectivo no qual exite a intervenção direta do pesquisador que aloca, de forma aleatória, o tratamento ao paciente.

Tipos de censura:

- Censura do tipo I: Ocorre em estudos em que, ao serem finalizados após um tempo pré-estabelicido, alguns indivíduos ainda não apresentaram o evento de interesse;
- Censura do tipo II: ocorre em estudos que são encerrados após a observação de um número pré-estabelicido de falhas;
- Censura aleatória: ocorre com frequência na área médica. Neste caso, o indivíduo saí do estudo sem a ocorrência do evento de interesse.

Seja T uma v.a. que representa o tempo de falha e C uma outra v.a. independente de T que representa o tempo de censura. O que se observa é:

$$t = min(T, C)$$

e:

$$\delta = 1, \text{se} T \leq C$$

$$\delta = 0, \text{se} T > C.$$

Ou seja:

$$\delta_i = I(T_i \leq C_i)$$

$$T \sim f_T(t|\theta_T)$$

$$C \sim f_C(c|\theta_C)$$

Suponha que os pares $(T_i, C_i), i = 1, 2, \ldots, n$ formam uma amostra aleatória de n indivíduos. Quando $C_1 = C$, uma constante fixada, obtém-se a censura do tipo I.

A função taxa de falha permite descrever a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t. É definida como:

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

A função de taxa de falha acumulada fornece a taxa de falha acumulada do indivíduo, sendo definida como:

$$\Lambda(u) = \int_0^t \lambda(u) du$$

O tempo de vida médio é dado por:

$$t_m = \mathbb{E}(T) = \int_0^\infty S(t) dt$$

A vida média residual mede, para indivíduos com tempo t, o tempo médio restante de vida e é dada por:

$$vmr(t) = \frac{\int_t^\infty S(u) du}{S(t)}$$

Em que $f(\cdot)$ é a densidade de T. Note que $vmr(0) = t_m$

Temos as seguintes relações:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{\delta}{\delta t}(\log S(t))$$

$$\Lambda(t) = \int_0^t \lambda(u)du = -log(S(t))$$

$$S(t) = \exp\{-\Lambda(t)\} = \exp\{-\int_0^t \lambda(u)du\}$$

$$S(t) = \frac{vmr(0)}{vmr(t)} \exp\{-\int_0^t \frac{du}{vmr(u)}\}$$

$$\lambda(t) = (\frac{\delta vmr(t)}{\delta t} + 1)/vmr(t)$$

Estimador de Kaplan-Meier

O estimador de Kaplan-Meier é uma adaptação da função de sobrevivência empírica que, na ausência de censuras, é definida como:

$$\hat{S}(t) = \frac{\text{nº de observações que não falharam até o tempo} t}{\text{nº de observações no estudo}}$$

- $\hat{S}(t)$ é uma função escada com degraus nos tempos observados de falha de tamanho $\frac{1}{n}$, em que n é o tamanho da amostra;
- Se existirem empates no tempo t, o tamanho do degrau fica multiplicado pelo número de empates;
- O EKM considera tanto intervalos de tempo quantos forem o número de falhas distrintas.

Construção do EKM em dados que envolvem censuras:

- Ordenar os tempos distintos de falha: $t_1 < t_2 < \ldots < t_k$;
- Obter d_j: número de falhas no tempo t_j;
- Obter n_j: número de observações sob risco (não falhou e não foi censurado) até o tempo t_j (exclusive);
- Obter $q_j = \frac{d_j}{n_j}$;
- a sobrevivência em t_j é estimada por:

$$\hat{S}(t) = (1-q_1)(1-q_2)\cdots(1-q_j) = \prod_{i\leq j}(1-\frac{d_i}{n_i})$$

O EKM é então definido por:

$$\hat{S}(t)\hat{S}(t) = \prod_{j:t_j<t}(\frac{n_j-d_j}{n_j}) = \prod_{j:t_j<t}(1-\frac{d_j}{n_j})$$

Principais propriedades:

- É não-viciado para amostras grandes;
- É fracamente consistente;
- Converge assintoticamente para um processo Gaussiano;
- É estimador de máxima verossimilhança de $S(t)$.

A variância assintótica do EKM é estimada pela fórmula de Greenwood:

$$\hat{Var}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{j:t_j<t} (\frac{d_j}{n_j(n_j - d_j)})$$

Como $\hat{S}(t)$, para t fixo, tem distribuição assintótica Normal, segue que um intervalo de confiança aproximado para $S(t)$ é dado por:

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

Existem correções para valores extremos de t, em que o intervalo acima pode apresentar valores menores que zero e maiores que 1. Uma solução é considerar alguma transformação para $S(t)$, por exemplo, para $\hat{U}(t) = \log[-\log \hat{S}(t)]$, temos:

$$\hat{Var}(\hat{U}(t)) = \frac{\sum_{j:t_j<t} \frac{d_j}{n_j(n_j-d_j)}}{[\sum_{j:t_j<t} \log(\frac{n_j-n_d}{n_j})]^2}$$

E o intervalo de confiança será dado por:

$$[\hat{S}(t)]^{\exp\{\pm z_{\alpha/2}\sqrt{\hat{Var}(\hat{U}(t))}\}}$$

O estimador de Nelson-Aalen se baseia na função de sobrevivência:

$$S(t) = \exp\{\Lambda(t)\}$$

Um estimador para $\Lambda(t)$ é:

$$\tilde{\Lambda}(t) = \sum_{j:t_k<t} (\frac{d_j}{n_j})$$

O estimador de Nelson-Aalen para $S(t)$ é dado por:

$$\tilde{S}(t) = \exp\{-\tilde{\Lambda}(t)\}$$

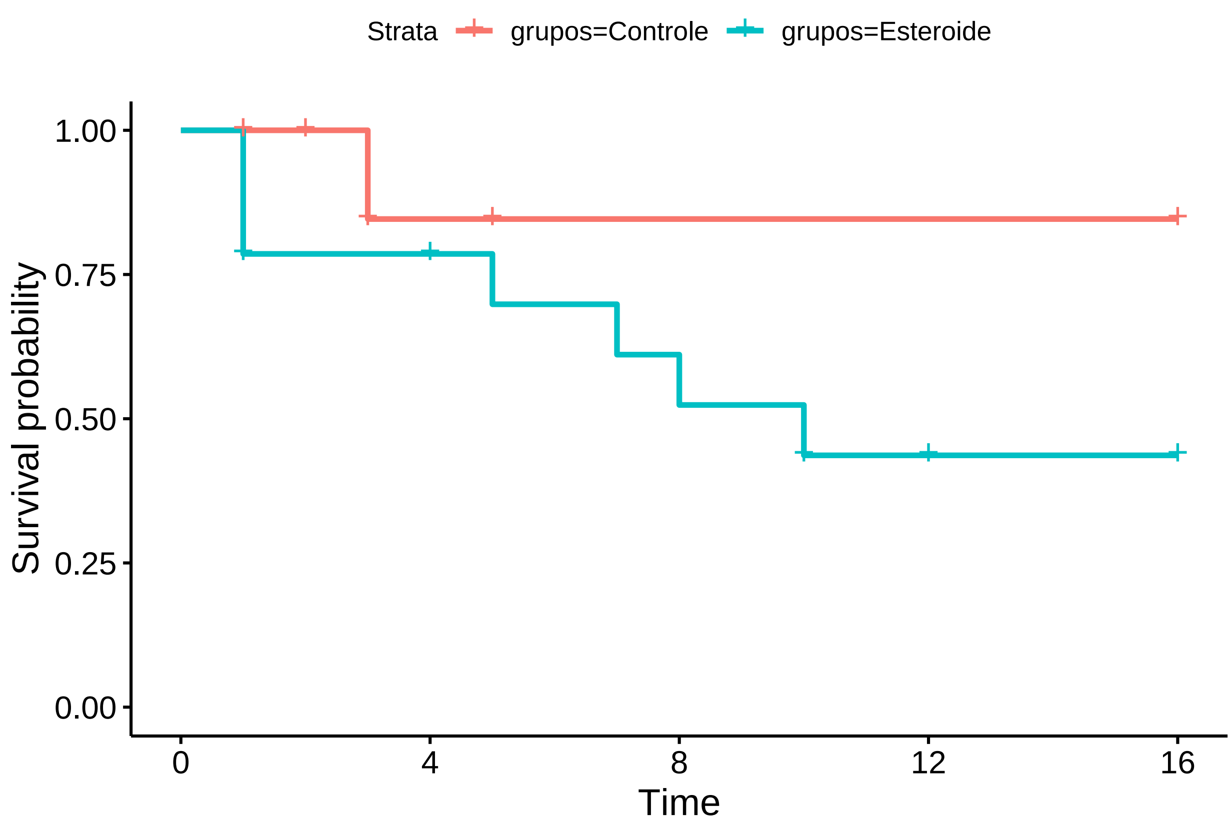
Com variância assintótica estimada:

$$\hat{Var}(\tilde{\Lambda}(t)) = [\tilde{S}(t)]^2 \sum_{j:t_k<t} (\frac{d_j}{n_j^2})$$

No R: Dados de Hepatite:

```
# Pacotes utilizados:
library(survival)
library(survminer)

# Dados dos tempos de falha:
tempos <- c(1,2,3,3,3,5,5,16,16,16,16,16,16,16,16,1,1,1,1,4,5,7,8,10,10,12,16,16,16)
```

```r
# Indicadora de censura:
cens <- c(0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,1,1,1,1,0,0,0,0,0)

# Divisão dos grupos:
grupos <- as.factor(c(rep("Controle", 15),rep("Esteroide", 14)))

# Dados completos:
dados <- data.frame(tempos, cens, grupos)

# Gráfico da curva de sobrevivência:
ekm <- survfit(Surv(tempos, cens) ~ grupos, data = dados)
ggsurvplot(ekm)
```




```r
# Sumários:
summary(ekm0 <- survfit(Surv(tempos, cens) ~ grupos, conf.type = "plain",data = dados))
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ grupos, data = dados,
##     conf.type = "plain")
##
##                 grupos=Controle
##         time       n.risk       n.event      survival       std.err lower 95% CI
##        3.000       13.000         2.000         0.846         0.100        0.650
## upper 95% CI
##        1.000
```

```
##
##                 grupos=Esteroide
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     14       3    0.786   0.110        0.571        1.000
##     5      9       1    0.698   0.128        0.448        0.948
##     7      8       1    0.611   0.138        0.340        0.882
##     8      7       1    0.524   0.143        0.243        0.805
##    10      6       1    0.437   0.144        0.155        0.718
```

```r
sob.NA <- survfit(coxph(Surv(tempos, cens) ~ grupos,data = subset(dados, grupos == "Esteroide"),method =
summary(sob.NA)
```

```
## Call: survfit(formula = coxph(Surv(tempos, cens) ~ grupos, data = subset(dados,
##     grupos == "Esteroide"), method = "breslow"))
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     14       3    0.807  0.0999        0.633        1.000
##     5      9       1    0.722  0.1201        0.521        1.000
##     7      8       1    0.637  0.1326        0.424        0.958
##     8      7       1    0.553  0.1394        0.337        0.906
##    10      6       1    0.468  0.1414        0.259        0.846
```

Estimação de Quantidades de Interesse:

- Probabilidade de sobrevivência
  - Estimador de Kaplan-Meier ou de Nelson-Aalen;
  - Interpolação pode ser útil;
  - Variância estimada pela fórmula de Greenwood. Transformações podem ser úteis.
- Percentis
  - Utilizar a inversa do estimador de Kaplan-Meier ou de Nelson-Aalen;
  - Interpolação pode ser útil;
  - Variância difícil de ser estimada.

A variância do estimador de percentis ($\hat{t}_p$) é dada por:

$$Var(\hat{t}_p) = \frac{V\hat{a}r(\hat{S}(\hat{t}_p))}{[f(\hat{t}_p)]^2}$$

Desejamos estimar:

$$t_m = \mathbb{E}(T) = \int_0^\infty S(t)dt$$

Uma estimativa paratmé substituir $S(t)$ por $\hat{S}(t)$. A integral se transformaem uma soma de áreas de retângulos, isto é:

$$\hat{t}_m = t_1 + \sum_{i=1}^{k-1} \hat{S}(t_j)(t_{j+1} - t_j)$$

Em que $t_1 < t_2 < \cdots < t_k$ são os k tempos de falha distintos e ordenados.