

# EPI10 - Análise de Sobrevivência

## Estimação da curva de sobrevivência (continuação)

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



## Estimadores de $S(t)$

## Estimadores de $S(t)$

- ▶ Dois outros estimadores para a função de sobrevivência  $S(t)$  são muito citados na literatura:
  - ▶ O estimador de Nelson-Aalen
  - ▶ O estimador atuarial (da tabela de vida)
- ▶ Não vamos estudá-los. Mas, faremos algumas considerações a respeito destes dois estimadores em comparação com o estimador de Kaplan-Meier.

## Estimadores de $S(t)$

- ▶ A grande diferença entre os estimadores de  $S(t)$  está no número de intervalos utilizados para a construção de cada um deles.
  - ▶ O estimador de Kaplan-Meier e o de Nelson-Aalen são sempre baseados em um número de intervalos igual ao número de tempos de falha distintos, enquanto que na tabela de vida, os tempos de falha são agrupados em intervalos de forma arbitrária.
  - ▶ Isto faz com que a estimativa obtida pelo estimador de Kaplan-Meier seja baseada frequentemente em um número de intervalos maior que a estimativa obtida através da tabela de vida.

## Estimadores de $S(t)$

- ▶ O uso da tabela de vida, considerando um número igual ou maior de intervalos que o do estimador de Kaplan-Meier, gera estimativas exatamente iguais às estimativas de Kaplan-Meier se o **mecanismo de censura** for do **tipo I** ou **II**.
- ▶ Entretanto, se o **mecanismo de censura** for do **tipo aleatório**, as estimativas serão próximas mas não necessariamente coincidentes.

## Estimadores de $S(t)$

- ▶ Alguns autores estudaram as propriedades assintóticas dos dois estimadores. Estes estudos mostraram a **superioridade do estimador de Kaplan-Meier**.
- ▶ Ele é um estimador não-viciado para a função de sobrevivência em grandes amostras, enquanto o estimador da tabela de vida não o é, com um vício que fica pequeno à medida que o comprimento dos intervalos diminuem.
- ▶ Com amostras de pequeno ou médio porte, existe alguma evidência empírica também da superioridade do estimador de Kaplan-Meier.
- ▶ Desta forma, o **mais indicado** é então usar o estimador de **Kaplan-Meier** ou eventualmente o de Nelson-Aalen, quando o interesse se concentrar em informações provenientes da função de sobrevivência.

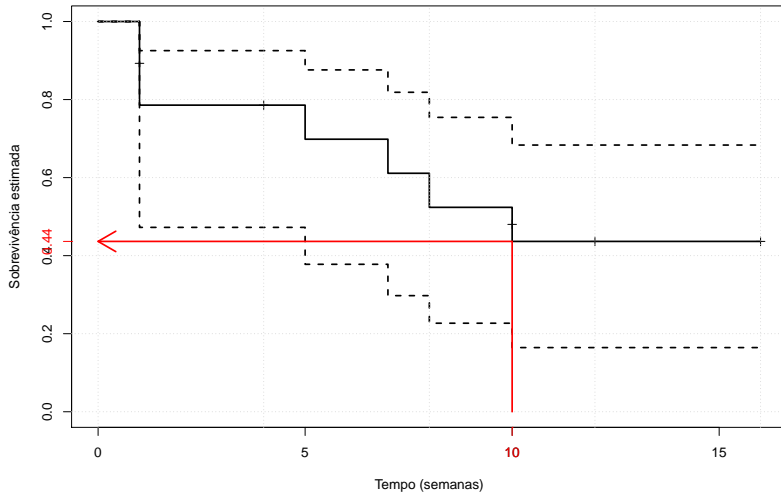
## Estimação de quantidades básicas

## Estimação de quantidades básicas

- ▶ A utilização direta da curva de Kaplan-Meier nos informa a probabilidade estimada de sobrevivência para um determinado tempo.
- ▶ Um exemplo é a probabilidade do indivíduo sobreviver a 10 semanas de tratamento. A estimativa de Kaplan-Meier para este valor é diretamente obtida do gráfico e é igual a 44%.



## Estimação de quantidades básicas



## Estimação de quantidades básicas

- ▶ Se o valor do tempo de interesse estiver ao longo de um degrau da curva de Kaplan-Meier, como  $t = 4$ , pode-se também utilizar

```
summary(ekm, times = 4)
```

```
## Call: survfit(formula = Surv(time = tempo, event = cens) ~ 1, data = df.hep,  
##      subset = grupo == "Esteroides", conf.type = "log-log")  
##
```

```
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI  
##      4      10        3   0.786   0.11      0.472      0.925
```

```
summary(ekm, times = c(4, 6, 7, 12))
```

```
## Call: survfit(formula = Surv(time = tempo, event = cens) ~ 1, data = df.hep,  
##      subset = grupo == "Esteroides", conf.type = "log-log")  
##
```

```
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI  
##      4      10        3   0.786   0.110      0.472      0.925  
##      6       8         1   0.698   0.128      0.378      0.876  
##      7       8         1   0.611   0.138      0.298      0.819  
##     12       4         2   0.437   0.144      0.164      0.683
```

## Estimação de quantidades básicas

- ▶ A partir da curva de Kaplan-Meier também é possível obter estimativas de percentis. Uma informação muito útil é o **tempo mediano de vida**.
- ▶ A mediana do tempo de sobrevivência é o tempo  $t_{0.5}$  que divide os tempos de sobrevivência em duas partes iguais:
  - ▶ 50% dos indivíduos com tempos de sobrevivência menores estão abaixo de  $\hat{t}_{0.5}$ ;
  - ▶ 50% dos indivíduos com maiores tempos de sobrevivência estão acima de  $\hat{t}_{0.5}$ .
- ▶ Ou seja,  $S(t_{0.5}) = 0.5$ .
  - ▶ No entanto, o estimador de Kaplan-Meier de  $S(t)$  é uma função escada, e geralmente não é possível obter uma estimativa para  $t_{0.5}$  tal que  $\hat{S}(\hat{t}_{0.5}) = 0.5$ <sup>1</sup>.

---

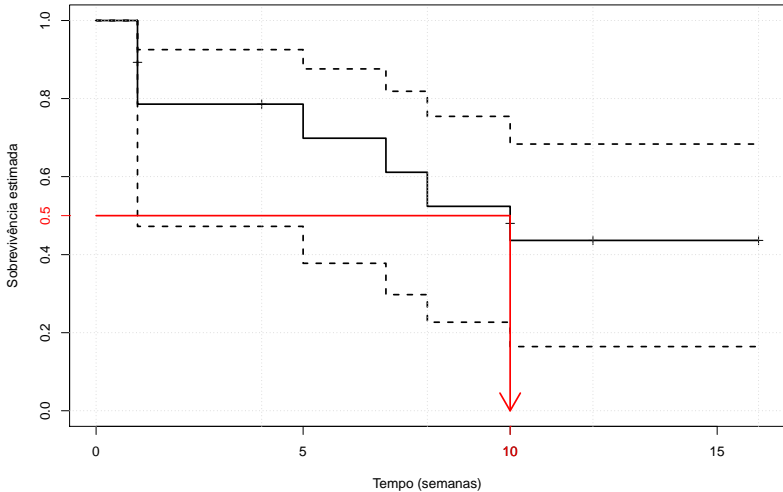
<sup>1</sup>Em verdade, podemos obter múltiplas estimativas.

## Estimação de quantidades básicas

- ▶ Assim, o estimador do tempo mediano,  $\hat{t}_{0.5}$  é definido como o **menor tempo de sobrevivência observado** tal que a estimativa da função  $S(t)$  é menor ou igual a 0.5, isto é,

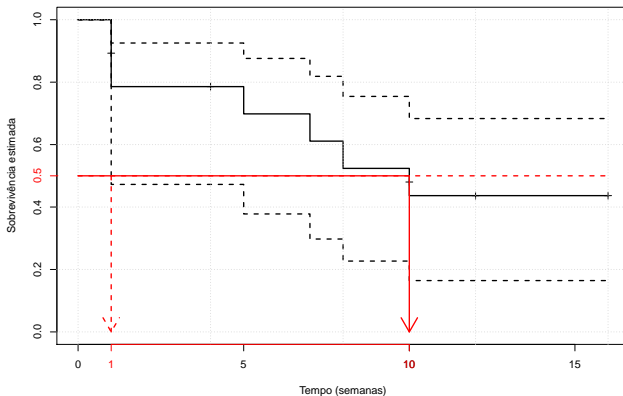
$$\hat{t}_{0.5} = \min \left\{ t_i \mid \hat{S}(t_i) \leq 0.5 \right\}.$$

## Estimación de quantidades básicas



## Estimação de quantidades básicas

- Uma solução análoga é obtida para estimar o intervalo de confiança do  $t_{0.5}$ .



## Estimação de quantidades básicas

- ▶ Em amostras pequenas, e/ou com poucos eventos, é possível que a mediana e/ou os limites do intervalo de confiança não sejam obtidos.
  - ▶ Algumas soluções alternativas foram propostas na literatura, tais como o uso de métodos de reamostragem (*bootstrap*).

```
ekm
```

```
## Call: survfit(formula = Surv(time = tempo, event = cens) ~ 1, data = df.hep,  
##      subset = grupo == "Esteroides", conf.type = "log-log")  
##  
##           n  events  median 0.95LCL 0.95UCL  
##          14      7      10        1      NA
```

## R: o pacote survminer



## R: o pacote survminer

- ▶ Como vimos anteriormente, o principal pacote em R para análise de sobrevivência é o `survival`.

```
library(survival)

ekm <- survfit(Surv(time = tempo, event = cens) ~ grupo,
               data = df.hep,
               conf.type = "log-log")
```

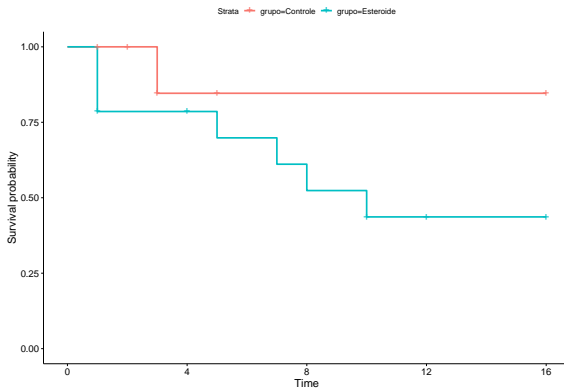
- ▶ Outros pacotes podem nos ajudar no ajuste de modelos e apresentação dos resultados. Este é o caso do pacote `survminer`.

## R: o pacote survminer

```
# install.packages("survminer")
```

```
library(survminer)
```

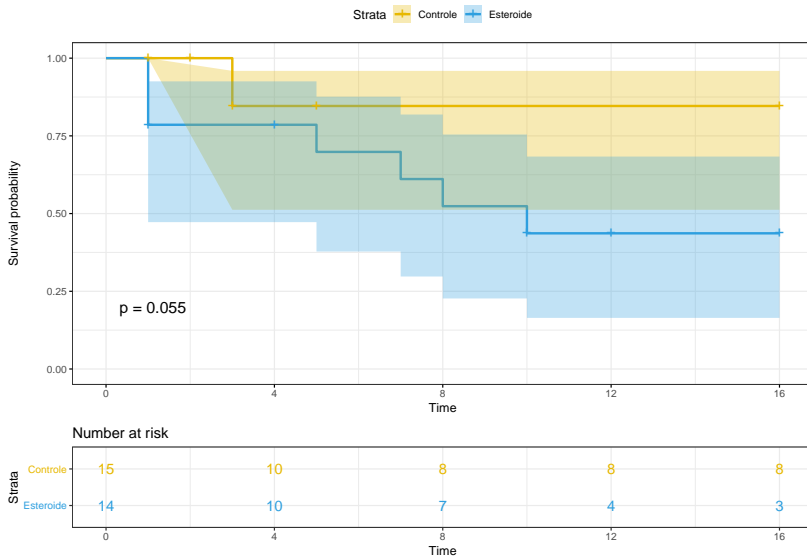
```
ggsurvplot(fit = ekm, data = df.hep)
```



## R: o pacote survminer

```
ggsurvplot(  
  fit = ekm,  
  data = df.hep,  
  size = 1,                # change line size  
  palette =  
    c("#E7B800", "#2E9FDF"), # custom color palettes  
  conf.int = TRUE,         # Add confidence interval  
  pval = TRUE,             # Add p-value  
  risk.table = TRUE,       # Add risk table  
  risk.table.col = "strata", # Risk table color by groups  
  legend.labs =  
    c("Controle", "Esteroides"), # Change legend labels  
  risk.table.height = 0.25, # Useful to change when you have multiple groups  
  ggtheme = theme_bw()     # Change ggplot2 theme  
)
```

# R: o pacote survminer



## Avisos

## Para casa

1. Repita os exemplos da aula.

# Próxima aula

- ▶ Comparação de curvas de sobrevivência.

# Por hoje é só!

Bons estudos!

