

# EPI10 - Análise de Sobrevida

## Conceitos básicos

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, 2021



# Introdução

# Introdução

- ▶ Em **análise de sobrevivência** a variável resposta é, geralmente, o **tempo até a ocorrência de um evento**<sup>1</sup> de interesse.
- ▶ De maneira mais geral, esta abrange um conjunto de técnicas estatísticas para a análise de variáveis aleatórias positivas, tais como
  - ▶ tempo até o óbito por determinada causa
  - ▶ tempo de permanência no hospital
  - ▶ tempo de duração de uma greve
  - ▶ tempo até o aprendizado de certa habilidade
  - ▶ medições de carga viral
  - ▶ o valor em reais que uma companhia de seguros paga a um determinado caso

---

<sup>1</sup>Também denominado tempo do evento, tempo de falha ou tempo de sobrevivência.

# Introdução

- ▶ Por exemplo, em estudos de câncer, é usual o registro das datas correspondentes ao diagnóstico da doença, à remissão, à recorrência da doença (recidiva) e óbito do paciente.
  - ▶ O tempo de falha pode ser, por exemplo, do diagnóstico até o óbito ou da remissão até a recidiva.

# Introdução

- ▶ A principal característica de dados de sobrevivência é a presença de **censura**<sup>2</sup>, que é a observação parcial da resposta.
- ▶ Isto se refere a situações em que, por alguma razão, o **acompanhamento** do paciente foi **interrompido**;
  - ▶ o participante mudou de cidade;
  - ▶ o estudo terminou para a análise dos dados;
  - ▶ o participante faleceu de causa diferente da estudada.

---

<sup>2</sup>Também chamada de observação censurada.

# Introdução

- ▶ Isto significa que toda informação referente à resposta se resume ao conhecimento de que **o tempo de falha é superior àquele observado**.
- ▶ Desta forma, faz-se necessário o uso dos **métodos de análise de sobrevivência**, que possibilitam incorporar na análise estatística a informação contida nos dados censurados.

## Caracterizando dados de sobrevivência

# Caracterizando dados de sobrevivência

- ▶ Os conjuntos de dados de sobrevivência são caracterizados pelos **tempos de falha** e, muito frequentemente, pelas **censuras**.
  - ▶ Estes dois componentes constituem a **variável resposta (desfecho)**.
- ▶ Em estudos clínicos, um **conjunto de covariáveis** é também, geralmente, medido em cada participante.
- ▶ Os seguintes três elementos constituem o tempo de falha:
  - ▶ o tempo inicial,
  - ▶ a escala de medida,
  - ▶ o evento de interesse (falha).



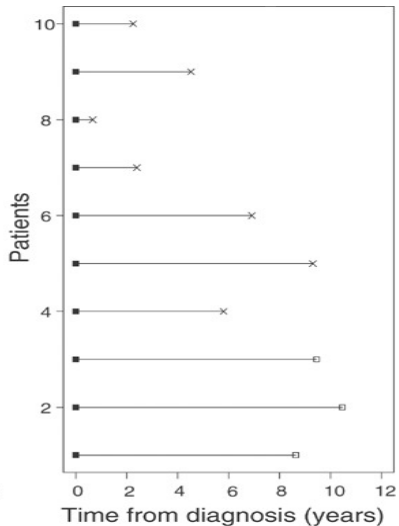
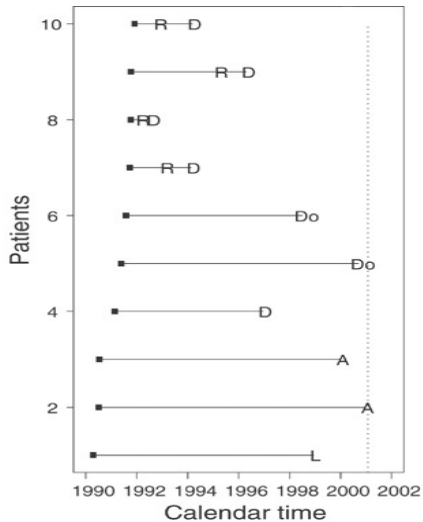
# Tempo de falha

- ▶ O **tempo de início** do estudo deve ser precisamente definido.
  - ▶ Em um estudo clínico aleatorizado, a data da aleatorização é a escolha natural para a origem do estudo.
  - ▶ A data do diagnóstico ou do início do tratamento de doenças também são outras escolhas possíveis.

# Tempo de falha

- ▶ A **escala de medida** é quase sempre o tempo real ou “de relógio”, apesar de existirem alternativas.
- ▶ Em testes de engenharia podem surgir outras escalas de medida, como o número de ciclos, a quilometragem de um carro ou qualquer outra medida de carga.

# Tempo de falha



# Tempo de falha

- ▶ O terceiro elemento é o **evento de interesse**.
- ▶ Estes eventos são, na maioria dos casos, indesejáveis e, como já mencionado, chamados de falha.
- ▶ É importante, em estudos de sobrevivência, definir de forma clara e precisa o que vem a ser a falha.
- ▶ **Ex:** o tempo de falha vai do tempo inicial de exposição (chegada ao supermercado) até o produto ficar “inapropriado ao consumo”.
  - ▶ Este evento deve ser claramente definido antes de iniciar o estudo.
  - ▶ Por exemplo, o produto fica inadequado para o consumo quando atingir mais que uma determinada concentração de microrganismos por  $mm^2$  de área do produto.

# Tempo de falha

- ▶ O evento de interesse (falha) pode ainda ocorrer devido a uma única causa ou devido a duas ou mais.
- ▶ Situações em que causas de falha competem entre si são denominadas na literatura de **riscos competitivos**.

# Censura

- ▶ Os estudos clínicos que envolvem uma resposta temporal são frequentemente prospectivos e de longa duração.
- ▶ Mesmo sendo longos, os estudos clínicos de sobrevivência usualmente terminam antes que todos os indivíduos no estudo venham a falhar.
- ▶ Uma característica decorrente destes estudos é, então, a presença de **observações incompletas** ou **parciais**.
- ▶ Estas observações, denominadas **censuras**, podem ocorrer por uma variedade de razões, dentre elas, a perda de acompanhamento do paciente no decorrer do estudo e a não ocorrência do evento de interesse até o término do experimento.

# Censura

- ▶ Note que toda informação obtida sobre estes indivíduos é que o seu tempo até o evento é **superior ao tempo registrado** até o último acompanhamento.
- ▶ Ressalta-se o fato de que, mesmo censurados, todos os resultados provenientes de um estudo de sobrevivência devem ser usados na análise estatística.
  - ▶ Mesmo sendo incompletas, as observações censuradas nos fornecem informações sobre o tempo de vida de pacientes;
  - ▶ A omissão das censuras no cálculo das estatísticas de interesse pode acarretar em conclusões enviesadas.

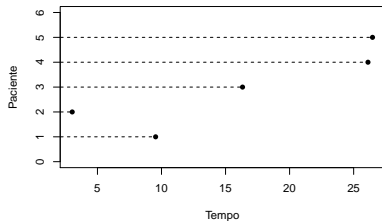
# Censura

- ▶ Os estudos apresentam diferentes **mecanismos de censura**.
- ▶ **Censura do tipo I** é aquela em que o estudo será encerrado após um período pré-estabelecido de tempo.
- ▶ **Censura do tipo II** é aquela em que o estudo será encerrado após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos.
- ▶ **Censura do tipo aleatório** é o que mais ocorre na prática médica/saúde.
  - ▶ Isto acontece quando um paciente/participante é retirado no decorrer do estudo sem ter ocorrido a falha.
  - ▶ Isto também ocorre, por exemplo, se o paciente/participante falecer por uma razão diferente da estudada.

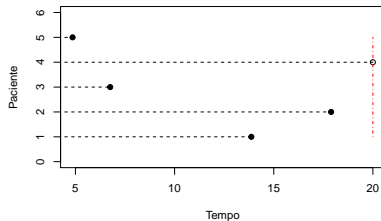


# Censura

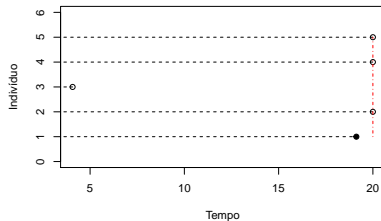
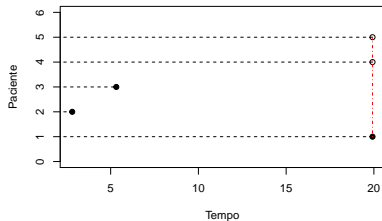
Dados completos



Dados com censura tipo I



Dados com censura tipo II



## Censura

- ▶ Uma representação simples do mecanismo de censura aleatória é feita usando duas variáveis aleatórias.
- ▶ Seja  $T$  uma variável aleatória representando o tempo de falha de um indivíduo e seja  $C$  uma variável aleatória independente de  $T$ , representando o tempo de censura associado a este indivíduo.
- ▶ Os dados observados são

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

e

$$\delta = \begin{cases} 1 & \text{se } T \leq C \\ 0 & \text{se } T > C. \end{cases}$$

# Censura

- ▶ Ainda temos a seguinte classificação das censuras.
- ▶ **Censura à direita:** o tempo de ocorrência do evento de interesse está à direita do tempo registrado.
- ▶ **Censura à esquerda:** o tempo registrado é maior que o tempo do evento.
  - ▶ **Exemplo:** um estudo para avaliar o tempo até a alfabetização. No começo do estudo alguns indivíduos já estão alfabetizados e não lembram em que tempo o evento ocorreu.
- ▶ **Censura intervalar:** o evento ocorreu em um intervalo de tempo em que se conhece apenas os limites do intervalo.
  - ▶ **Exemplo:** um veterinário realiza visitas a cada três meses em casas de uma certa vizinhança para acompanhar a saúde de cães. O evento de interesse é a soroconversão para leishmaniose que ocorre entre as visitas.

# Censura

- ▶ Em nosso curso, a atenção está voltada aos dados de sobrevivência com censura à direita, que é a situação encontrada com mais frequência em estudos epidemiológicos.
- ▶ Desta forma, quando for simplesmente mencionada a palavra **censura** entenda-se **censura à direita**.

# Representação dos dados de sobrevivência

## Representação dos dados de sobrevivência

- ▶ Os dados de sobrevivência para o indivíduo  $i$  ( $i = 1, \dots, n$ ) sob estudo, são representados, em geral, pelo par  $(t_i, \delta_i)$  sendo  $t_i$  o **tempo de falha ou de censura** e  $\delta_i$  a variável indicadora de falha ou censura, isto é,

$$\delta_i = \begin{cases} 1 & \text{se } t_i \text{ é um tempo de falha} \\ 0 & \text{se } t_i \text{ é um tempo de censura.} \end{cases}$$

- ▶ Na presença de covariáveis medidas no  $i$ -ésimo indivíduo tais como  $x_i = (\text{sexo}_i, \text{idade}_i, \text{tratamento}_i)$ , os dados ficam representados por  $(t_i, \delta_i, x_i)$ .
- ▶ No caso especial de dados de sobrevivência intervalar tem-se, ainda, a representação  $(\ell_i, u_i, \delta_i, x_i)$  em que  $\ell_i$  e  $u_i$  são, respectivamente, os limites inferior e superior do intervalo observado para o  $i$ -ésimo indivíduo.

## Exemplo

## Estudo de Hepatite

- ▶ Um estudo clínico aleatorizado foi realizado para investigar o efeito da terapia com esteroide no tratamento de hepatite viral aguda.
- ▶ **Vinte e nove pacientes com esta doença** foram aleatorizados para receber um placebo ou o tratamento com esteroide.
- ▶ Cada paciente foi acompanhado por 16 semanas ou até o óbito (evento de interesse) ou até a perda de acompanhamento.



# Estudo de Hepatite

- ▶ Os tempos de sobrevivência observados, em semanas, para os dois grupos são apresentados na tabela a seguir (+ indica censura).

| Grupo     | Tempo de sobrevivência em semanas                                |
|-----------|--|
| Controle  | 1+, 2+, 3, 3, 3+, 5+, 5+, 16+, 16+, 16+, 16+, 16+, 16+, 16+, 16+ |
| Esteróide | 1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+            |

## Especificando o tempo de sobrevivência

# Especificando o tempo de sobrevivência

- ▶ A variável aleatória não negativa  $T$ , que representa o tempo de falha, é usualmente especificada em análise de sobrevivência pela sua **função de sobrevivência** ou pela **função de taxa de falha (ou risco)**.

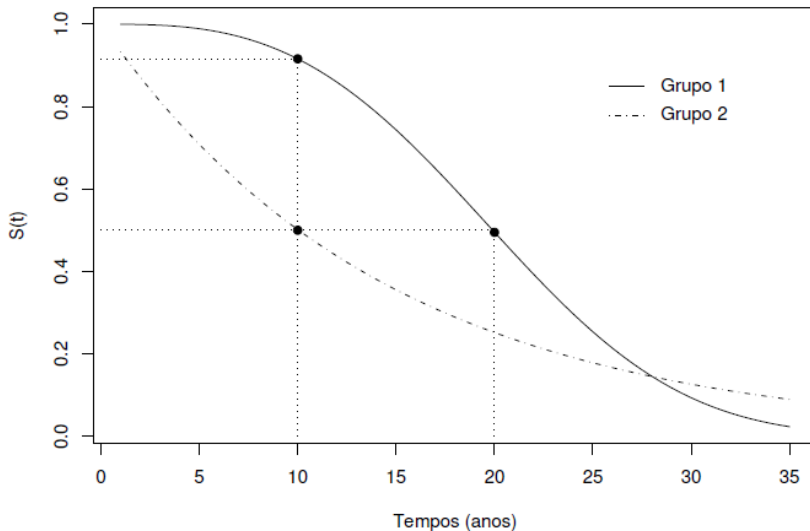
# Função de sobrevivência

- ▶ A função de sobrevivência é definida como a **probabilidade de uma observação não falhar até um certo tempo  $t$** , ou seja, a **probabilidade de uma observação sobreviver ao tempo  $t$** .
- ▶ Em termos probabilísticos, isto é escrito como

$$S(t) = \Pr(T > t).$$

- ▶ Logo, a função de distribuição de probabilidade pode ser expressa como  $F(t) = 1 - S(t)$  (probabilidade de uma observação não sobreviver ao tempo  $t$ ).

# Função de sobrevivência



## Função de taxa de falha ou de risco

- ▶ A probabilidade da falha ocorrer em um intervalo de tempo  $[t_1, t_2)$  pode ser expressa em termos da função de sobrevivência como

$$S(t_1) - S(t_2).$$

## Função de taxa de falha ou de risco

- ▶ A **taxa de falha no intervalo**  $[t_1, t_2)$  é definida como a probabilidade de que a falha ocorra neste intervalo, dado que não ocorreu antes de  $t_1$ , dividida pelo comprimento do intervalo.
- ▶ Assim, a taxa de falha no intervalo  $[t_1, t_2)$  é expressa por

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)}.$$

## Função de taxa de falha ou de risco

- De uma forma geral, redefinindo o intervalo como  $[t, t + \Delta t)$ , a expressão anterior assume a seguinte forma

$$\lambda(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}.$$



## Função de taxa de falha ou de risco

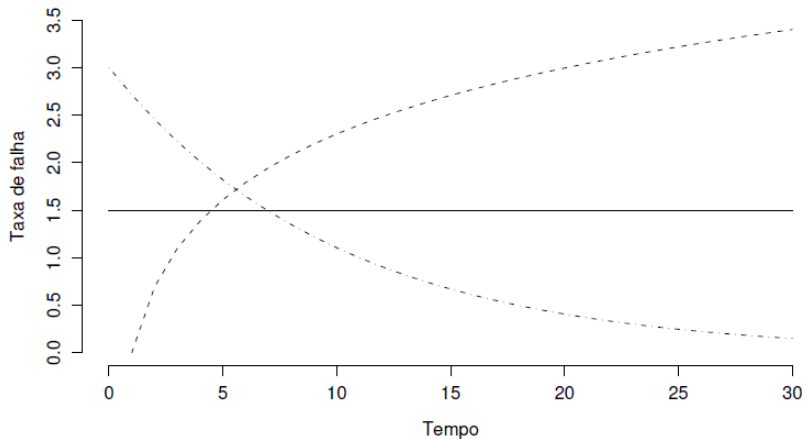
- ▶ Assumindo  $\Delta t$  bem pequeno,  $\lambda(t)$  representa a **taxa de falha instantânea no tempo  $t$  condicional à sobrevivência até o tempo  $t$** .
- ▶ Observe que as taxas de falha são números positivos, mas sem limite superior.
- ▶ A função de taxa de falha  $\lambda(t)$  é bastante útil para descrever a distribuição do tempo de vida de pacientes/participante.
- ▶ Ela descreve a forma em que a taxa instantânea de falha muda com o tempo.

## Função de taxa de falha ou de risco

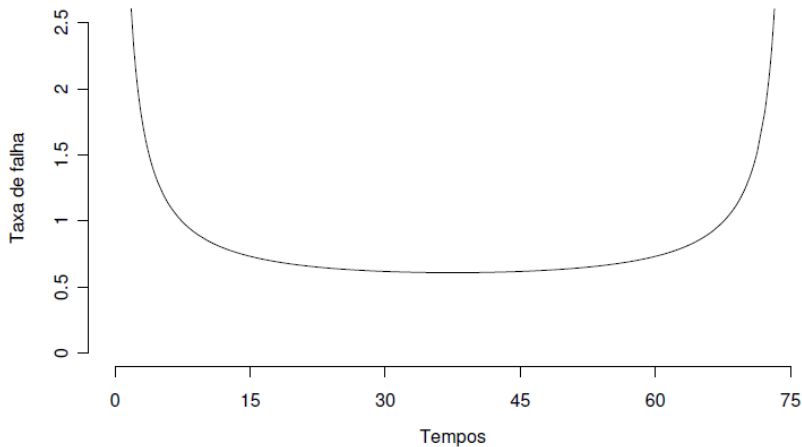
- A função de taxa de falha de  $T$  é então definida como

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

## Função de taxa de falha ou de risco



## Curva da banheira



## Função de taxa de falha acumulada

- ▶ Outra função útil em análise de sobrevivência é a **função de taxa de falha acumulada**.
- ▶ Esta função fornece a taxa de falha acumulada do indivíduo é definida por:

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

## Relações entre as funções

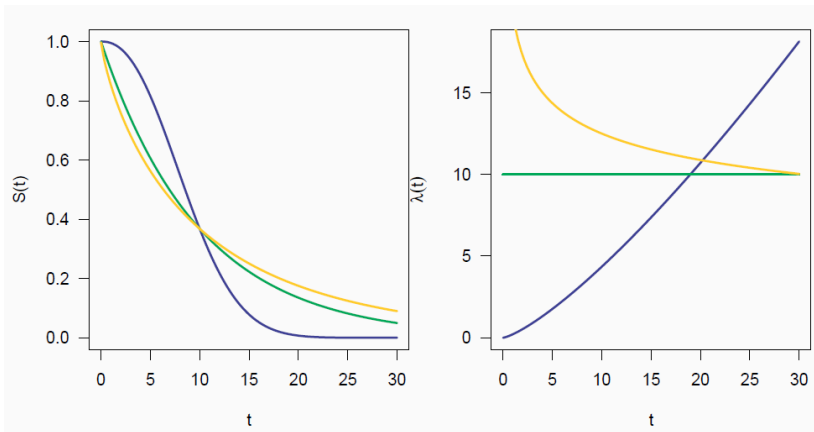
- Para  $T$  uma variável aleatória contínua não negativa, tem-se, em termos das funções definidas anteriormente, algumas relações matemáticas importantes entre elas.

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \{\log[S(t)]\},$$

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

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

## Relações entre as funções



## Para casa

1. Leia o capítulos 1 do livro **Análise de sobrevivência aplicada**<sup>3</sup>.
2. Leia os capítulos 1, 2 e 3 do livro **Análise de sobrevivência: teoria e aplicações em saúde**<sup>4</sup>.
3. Resolver os exercícios 1 e 2 do Capítulo 2, 1 e 2 do Capítulo 3 do livro **Análise de sobrevivência: teoria e aplicações em saúde**.

---

<sup>3</sup>Colosimo, E. A. e Giolo, S. R. **Análise de sobrevivência aplicada**, Blucher, 2006.

<sup>4</sup>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 e comparação de curvas de sobrevivência.

# Por hoje é só!

Bons estudos!

