

## CONCEITOS E ABORDAGEM EM ANÁLISE DE SOBREVIVÊNCIA/CONFIABILIDADE

**Autor:** Romulo Barros de Freitas

**Orientador:** Manoel Ferreira dos Santos Neto

**Tutor:** Ronald Targino Nojosa

### 1. Introdução

A análise de sobrevivência, também conhecida como análise de confiabilidade, possui diversos campos de aplicação, como medicina, engenharia e ciências sociais. De acordo com Colosimo e Giolo [2, pág. 6], os dados de sobrevivência/confiabilidade são caracterizados pelos tempos de falha e, muitas vezes, pelas censuras. Estes dois componentes constituem a resposta de interesse na análise.

### 2. Objetivos

Objetiva-se, com este trabalho, apresentar alguns conceitos e técnicas estatísticas utilizadas na análise de sobrevivência, bem como a sua aplicação em um conjunto de dados relacionados ao câncer de pulmão, com a finalidade de ilustrar, de forma breve e sucinta a importância da área para a sociedade.

### 3. Metodologia

O trabalho foi feito por meio de uma pesquisa bibliográfica, e como ferramenta computacional foi utilizado o *software* R [3], gratuito e de código aberto.

#### 3.1 Representação dos Dados de Sobrevida

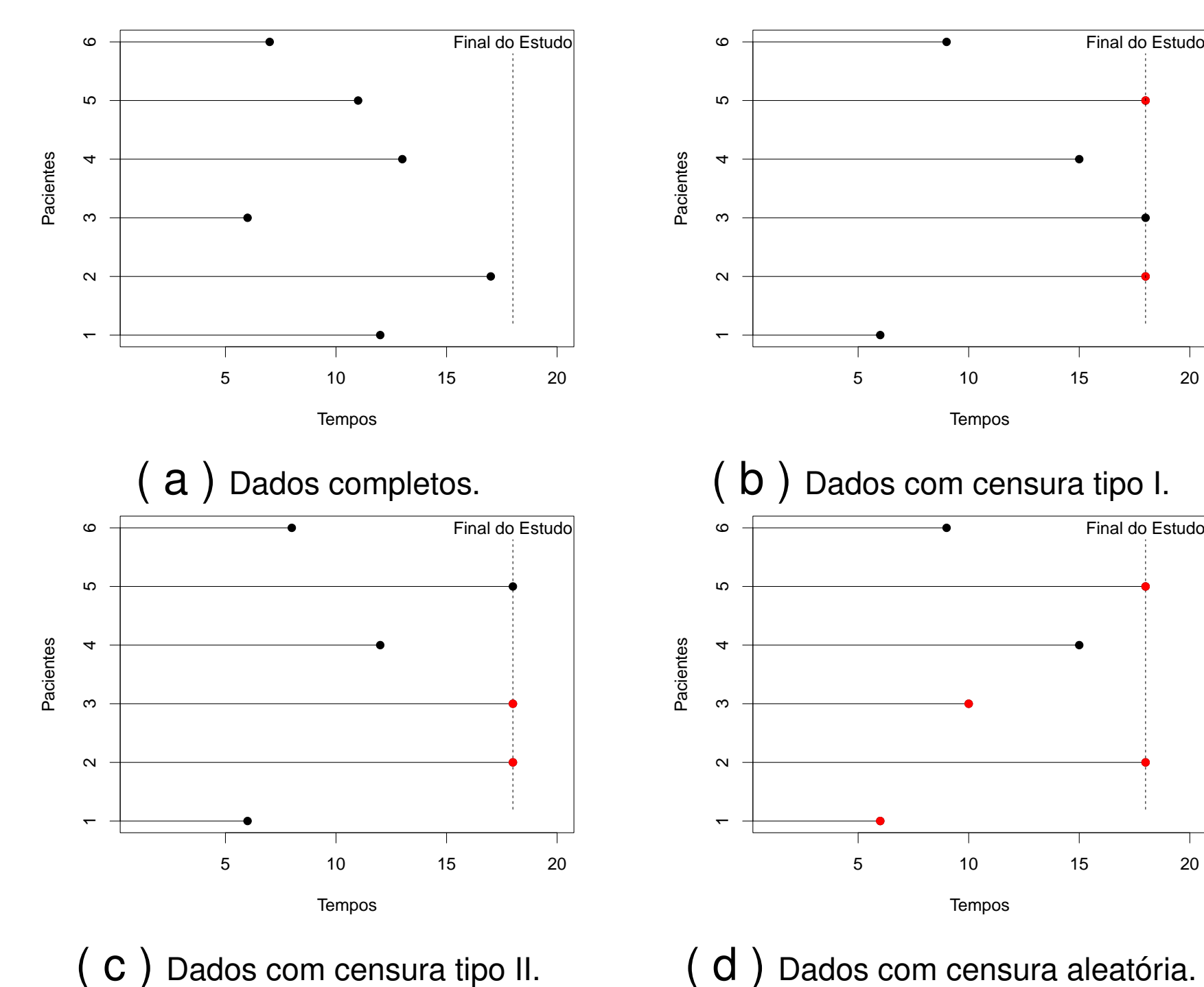
Para cada indivíduo  $i$ , ( $i = 1, 2, \dots, n$ ) no estudo de interesse, a representação dos dados de sobrevivência é dada pelo par  $(t_i, \delta_i)$ , em que  $t_i$  é o tempo de falha ou de censura, e  $\delta_i$  a variável indicadora de falha ou de censura, ou seja,

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

#### 3.2 Tipos de Censura

Na Figura 1 são apresentados alguns mecanismos de censura à direita, situação que é frequentemente encontrada em estudos de sobrevivência.

Figura 1: Gráficos representando alguns mecanismos de censura.



- Censura do tipo I:** quando o estudo termina após um período pré-estabelecido de tempo. As observações cujo evento de interesse não foi observado até este tempo são ditas censuradas;
- Censura do tipo II:** quando o estudo termina após ter ocorrido o evento de interesse para um número pré-estabelecido de observações;
- Censura aleatória:** quando a observação for

retirada no decorrer do estudo sem ter ocorrido o evento de interesse ou o evento de interesse tenha ocorrido por uma razão diferente da estudada.

#### 3.3 Função de Sobrevida

A função de sobrevivência é dada por:

$$S(t) = \Pr(T > t) = 1 - \Pr(T \leq t),$$

em que  $\Pr(T > t)$  é a probabilidade de  $T$  superar o valor  $t$ .

##### 3.3.1 Propriedades de $S(t)$

- $S(t)$  é uma função sempre não crescente;
- $S(t)$  é uma função contínua (no caso discreto sempre contínua à direita);
- Conforme a variável tempo aumenta, a função sobrevivência tende a 0 e vice e versa;
- $\Pr(t_1 \leq T \leq t_2) = S(t_1) - S(t_2)$ .

#### 3.4 Função Taxa de Falha

A função taxa de falha é definida por:

$$\lambda(t) = \frac{f(t)}{S(t)},$$

onde,  $f(t)$  é a função densidade de probabilidade e  $S(t)$  é a função de sobrevivência.

#### 3.5 Estimador de Kaplan-Meier

O estimador de Kaplan-Meier, denotado por  $\hat{S}_{km}(t)$ , é utilizado para estimar a função de sobrevivência e é definido por:

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

em que  $t_j$  são os tempos em que ocorreram as falhas na amostra,  $d_j$  é o número total de falhas ocorridos precisamente em  $t_j$  e  $n_j$  é o número total de falhas ocorridos precisamente em  $t_j$ .

#### 3.6 Tempo Total em Teste (Curva TTT)

De acordo com Aarset [1], se a curva TTT apresentar um formato convexo ou côncavo, a função taxa de falha será monotonicamente crescente ou monotonicamente decrescente, respectivamente, sendo assim, os possíveis modelos probabilísticos associados ao tempo de sobrevivência serão exponencial, gama e Weibull. Outras opções, são os modelos log-logística e log-normal no caso da curva TTT apresentar um formato côncavo e então convexo, indicando uma taxa de falha unimodal. É importante destacar que no trabalho foi utilizado a versão da curva TTT de quando há dados censurados, apresentada por Westberg e Klefsjö [4], a qual se utiliza do estimador de Kaplan-Meier.

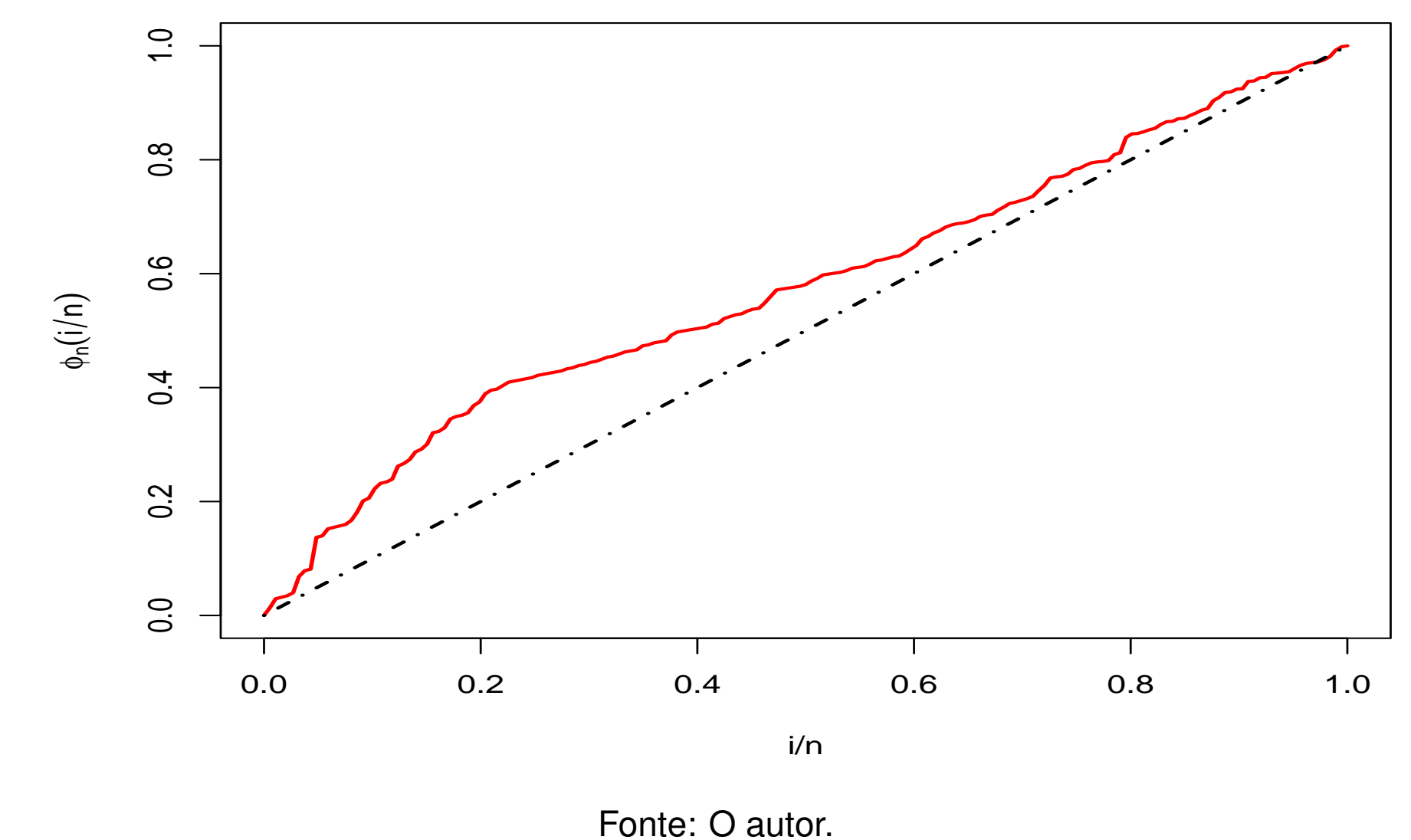
### 4. Aplicação

O conjunto de dados analisado refere-se a um estudo realizado pela empresa *North Central Cancer Treatment Group* com pacientes com câncer de pulmão avançado, o qual está disponível para análise no *software* R por meio do pacote *survival*.

Note na Figura 2 que a curva TTT é côncava, o que indica que a função de risco é crescente (o gráfico foi construído por meio do pacote *EstimationTools*). Desta maneira, pode-se utilizar algum

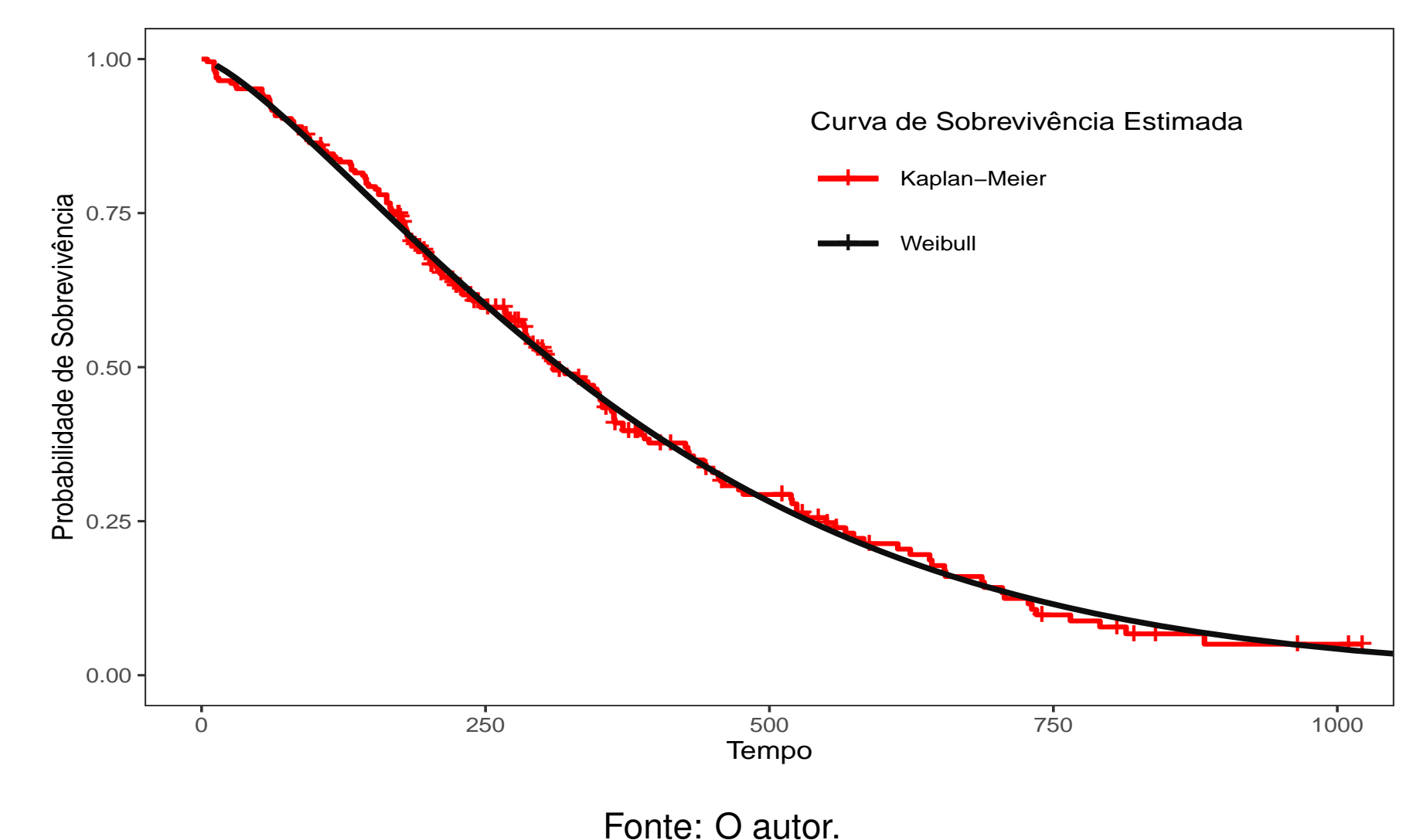
modelo que tenha esta característica. Aqui será utilizado o modelo Weibull para ajustar os dados.

Figura 2: Curva TTT para o conjunto de dados de câncer de pulmão.



Na Figura 3, é apresentado o resultado do ajuste do modelo Weibull aos dados. Note que a função de sobrevivência estimada pelo estimador de Kaplan-Meier, e a função de sobrevivência estimada pelo modelo Weibull são bem próximas, o que indica um bom ajuste. Para a construção do gráfico foi utilizada a função *ggsurvplot* do pacote *survminer*.

Figura 3: Curvas de sobrevivência estimadas para o conjunto de dados de câncer de pulmão.



### 5. Conclusões

Neste trabalho foram apresentados alguns dos principais conceitos que compõem a análise de sobrevivência, como as características dos dados, a função de sobrevivência/confiabilidade e suas propriedades, o estimador de Kaplan-Meier e a curva TTT. Além disso, foi realizado o ajuste de um conjunto de dados real referente a um estudo sobre câncer de pulmão. No ajuste observamos que a distribuição Weibull é uma boa candidata para modelar os dados.

### Referências

- [1] Magne Vollen Aarset. "How to Identify a Bathtub Hazard Rate". Em: *IEEE Transactions on Reliability* R-36.1 (abr. de 1987), pp. 106–108. ISSN: 1558-1721. DOI: 10.1109/TR.1987.5222310. URL: <https://ieeexplore.ieee.org/document/5222310/> (acesso em 16/10/2023).
- [2] Enrico Antônio Colosimo e Suely Ruiz Giolo. *Análise de sobrevivência aplicada*. por. 1. ed. ABE - Projeto Fisher. OCLC: 230883057. São Paulo: Edgard Blücher, 2006. ISBN: 9788521203841.
- [3] *Download R-4.3.1 for Windows. The R-project for statistical computing*. URL: <https://cran.r-project.org/bin/windows/base/> (acesso em 25/10/2023).
- [4] Ulf Westberg e Bengt Klefsjö. "TTT-Plotting for Censored Data Based on the Piecewise Exponential Estimator". en. Em: *International Journal of Reliability, Quality and Safety Engineering* 01.01 (mar. de 1994), pp. 1–13. ISSN: 0218-5393, 1793-6446. DOI: 10.1142/S0218539394000027. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0218539394000027> (acesso em 30/10/2023).