UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
DEPARTAMENTO DE ESTATÍSTICA
ANÁLISE DE SOBREVIVÊNCIA

Relatório abordagem Análise de Sobrevivência

Aluna: Vitória Nascimento de Jesus Sesana

Professora: Agatha Sacramento Rodrigues

Vitória - ES
2024

Sumário

1 Introdução **2**

2 Metodologia **2**

2.1 Função de Sobrevivência 2

2.2 Estimador Kaplan-Meier 2

2.3 Teste de Log-Rank 2

2.4 Modelos probabilísticos 3

3 Cenário 1: Doença de Hodkings **3**

3.1 Análise Descritiva 3

3.2 Aplicações Não-Paramétricas 3

3.2.1 Kaplan-Meier por Gêneros 3

3.2.2 Kaplan-Meier por Faixas Etárias 3

3.2.3 Kaplan-Meier por Tipos de Linfoma 3

3.2.4 Kaplan-Meier por Estágios da Doença 3

3.3 Conclusões 3

4 Cenário 2: Resistência de Isolentes Térmicos **4**

4.1 Análise Descritiva 4

4.2 Modelagens Paramétricas 4

4.3 Estimativas 4

4.3.1 Tempo Mediano 4

4.3.2 Tempo Médio 4

4.3.3 Sobrevida de 80 % 4

4.4 Conclusões 4

5 Conclusões Finais **4**

6 Referências **5**

7 Apêndices **6**

7.1 Códigos utilizados no Cnário 1 6

7.2 Códigos utilizados no Cnário 2 7

1 Introdução

A necessidade de analisar o tempo até a ocorrência de um evento é demandado em diversas áreas do conhecimento. Um exemplo, aplicado na área da saúde, seria o interesse em verificar o tempo até o óbtito de pacientes após receberem um tratamento específico para uma doença grave, como câncer.

No entanto, é comum no processo da coleta dos dados haja unidades com imprecisão ou a não observação do tempo exato da ocorrência do evento. Diz-se que estas unidades são unidades censuradas ou que sofreram censura.

Intuitivamente, muitos optam por eliminar as observações que contém censura, mesmo com algum tipo de informação já coletada. Desse modo, a perda de informação impactaria nas conclusões finais relacionadas ao estudo. Diante disso, a Análise de Sobrevivência busca solucionar esse problema com um conjunto de técnicas que permitem incluir a censura em seus cálculos.

A análise de dobrevivência pode ser abordada por duas formas: paramétrica e não-paramétrica. Os métodos paramétricos, ao contrário dos métodos não-paramétricos, pressupõem que os dados são provenientes de um tipo de distribuição de probabilidade. Cada metodologia possui suas prórpias técnicas a fim de concluir informações a respeito do problema em estudo.

O objetivo desse relatório é aplicar tais abordagens em dois cenários distintos. O primeiro cenário é relacionado ao tempo de vida de pacientes diagnosticados com a doença de Hodgkings, utilizando métodos não-paramétricos. Em contraponto, o segundo cenário possui como foco o tempo de resistência de isolantes térmicos até a sua ruptura, quando submetidos a altas tensões. Neste caso, será utilizado métodos paramétricos.

2 Metodologia

2.1 Função de Sobrevivência

Seja uma variável aleatória não negativa T, usualmente contínua, que mede o tempo até a ocorrência de um evento, também chamada como tempo de falha. A função sobrevivência indica a probabilidade da observação não falhar após o tempo t.

$$
\begin{aligned}
S(t) &= P(T \geq t) \\
&= 1 - P(T \leq t) \\
&= 1 - F(t)
\end{aligned} \tag{1}
$$

O objetivo geral é estimar a função de sobrevivência.

2.2 Estimador Kaplan-Meier

O estimador de Kaplan-Meier, ou estimador limite-produto, é um método não-paramétrico utilizado para estimar a função de sobrevivência $S(t)$. Uma de suas vantagens é a não necessidade de supor distribuição probabílistica aos dados, ao mesmo tempo que considera as censuras em seus cálculos. No caso sem censura, o estimador Kaplan-Meier é a própria função de sobrevivência empírica:

$$
\widehat{S(t)} = \frac{\text{n}^{\text{o}} \text{ de observações que não falharam até o tempo t}}{\text{n}^{\text{o}} \text{ total de observações no estudo}} \tag{2}
$$

$$
\begin{aligned}
\widehat{S(t)} &= \prod_{j:t_j<t} \\
&= 1 - F(t)
\end{aligned} \tag{3}
$$

2.3 Teste de Log-Rank

$$
\begin{aligned}
&H_0 = S_1(t) = \cdots = S_j(t); \forall t\epsilon \\
&H_1 = \epsilon
\end{aligned} \tag{4}
$$

2.4 Modelos probabilísticos

exponencial weibull log normal log logisticos

comparar S(t) klapan meier com os S(t) do modelo Estimação Para cada modelo probabilístico há um ou mais parâmetros que caracterizam seu comportamento para adequar ao contexto do problema. Esses valores são desconhecidos e para obtê-los são utilizadas técnicas estatísticas inferenciais. Há vários métodos de estimação, um dos mais convencionais é o método de mínimos quadrados. Entretanto, este desconsidera as censuras presentes. Desse modo, uma opção é utilizar o Método da Máxima Verossimilhança, pois este incorpora a censura em seus cálculos.

Computacional R e Latex

3 Cenário 1: Doença de Hodkings

Segundo o Ministério da Saúde, o linfoma ou doença de Hodkings é um câncer que afeta o sistema linfático, conjunto de tecidos e órgãos responsáveis pela imunidade e vasos que conduzem essas células através do corpo.

A partir disso, Bartolucci e Fraser (1977) apresentaram uma base de dados de 60 pessoas que receberam tratamento padrão para a doença com as seguintes variáveis:

- **id**: Identificação do paciente;
- **survivaltime**: Tempo de vida, em meses;
- **dead**: Indicador de falha/morte (0 - não; 1 - sim);
- **age**: Idade em anos;
- **sex**: Gênero (0 - feminino; 1- masculino);
- **stage**: Estágio da doença (0 - inicial; 1 - avançado);
- **hist**: Tipo de linfoma (1 - esclerose nodular; 2 - misto celular; 3 - depleção de linfócitos).

Neste cenário, a variável de interesse é o tempo de vida, em meses, dos pacientes com a doença de Hodkings (survivaltime) e a variável indicadora dos, as demais são covariáveis.

Será calculado as estimativas pelo método de Kaplan-Meier nos níveis das covariáveis *sex*, *age*, *hist* e *stage* a fim de analisar o comportamento do tempo de sobrevida dos pacientes dos grupos. Não somente, também será aplicado o teste de Log-Rank para testar a proporcionalidade entre os níveis de cada covariável.

Vale ressaltar que a covariável *age* foi categorizada para se adequar a solução proposta.

- **age**: Idade intervalar ("53 anos ou mais"; "38 anos até menos 53 anos"; "25 anos até menos 38 anos"; "menos de 25 anos").

3.1 Análise Descritiva

3.2 Aplicações Não-Paramétricas

3.2.1 Kaplan-Meier por Gêneros

A Figura 1

3.2.2 Kaplan-Meier por Faixas Etárias

3.2.3 Kaplan-Meier por Tipos de Linfoma

3.2.4 Kaplan-Meier por Estágios da Doença

3.3 Conclusões

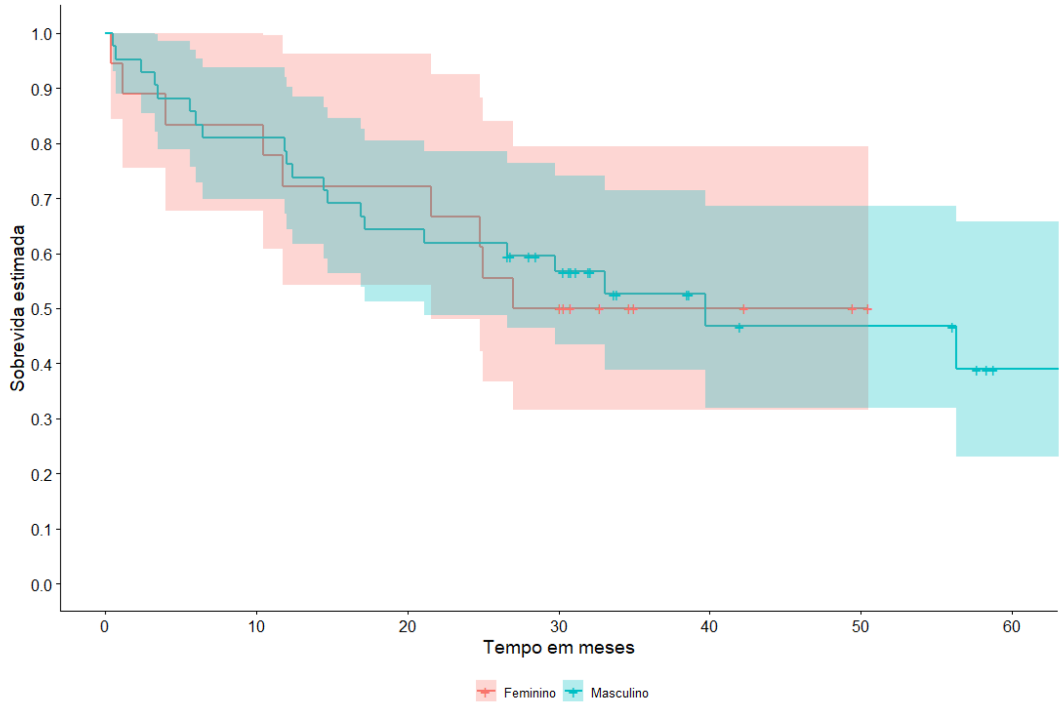

Figura 1: Estimativas por gênero da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

4 Cenário 2: Resistência de Isolentes Térmicos

0.19	0.78	0.96	1.31	2.78	3.16	4.67	4.85
6.50	7.35	8.27	12.07	32.52	33.91	36.71	

4.1 Análise Descritiva

4.2 Modelagens Paramétricas

4.3 Estimativas

Os resultados a seguir adotaram o modelo selecionado com base nos tópicos anteriores.

4.3.1 Tempo Mediano

4.3.2 Tempo Médio

4.3.3 Sobrevida de 80 %

4.4 Conclusões

5 Conclusões Finais

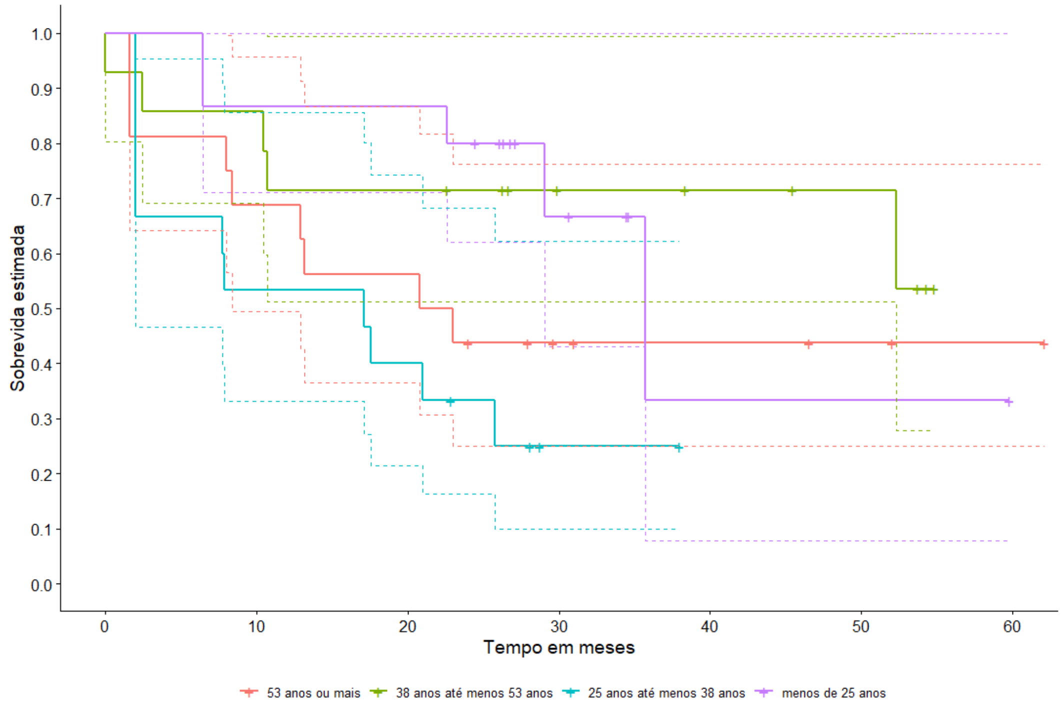

Figura 2: Estimativas por faixa etária da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

6 Referências

1. COLOSIMO, Enrico Antônio; GIOLO, Suely. Análise de sobrevivência aplicada. São Paulo: Blucher, c2006. 367p

2. Bartolucci, A. A., & Fraser, M. D. (1977). Comparative Step-Up and Composite Tests for Selecting Prognostic Indicators Associated with Survival. Biometrical Journal, 19(6), 437–448. doi:10.1002/bimj.4710190607

3. BRASIL. Ministério da Saúde. Instituto Nacional de Câncer - INCA. Linfoma de Hodgkin, 2022. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/linfoma-de-hodgkin. Acesso em: 01 mai. 2024.

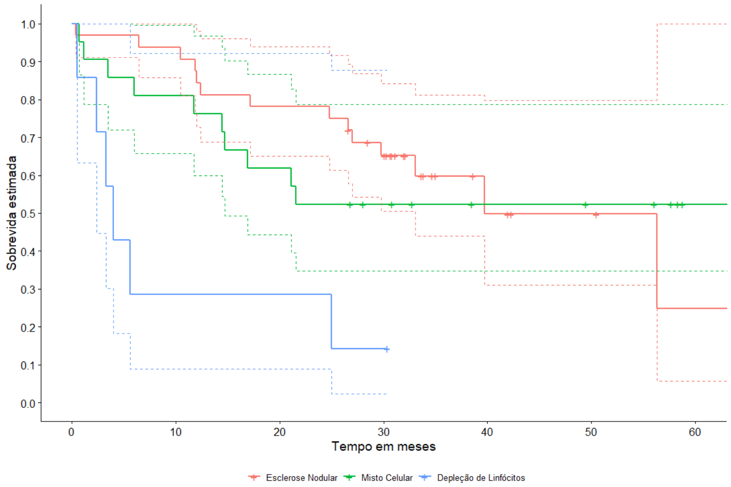

Figura 3: Estimativas por tipo de linfoma da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

7 Apêndices

7.1 Códigos utilizados no Cnário 1

```
require(survival)

tempos<-c(3,5,6,7,8,9,10,10,12,15,15,18,19,20,22,25,28,30,40,45)
cens<-c(1,1,1,1,1,1,1,0,1,1,0,1,1,1,1,1,1,1,1,0)

ajust1<-survreg(Surv(tempos,cens)~1,dist="exponential")
ajust1
alpha<-exp(ajust1$coefficients[1])
alpha

ajust2<-survreg(Surv(tempos,cens)~1,dist="weibull")
ajust2

alpha<-exp(ajust2$coefficients[1])

gama<-1/ajust2$scale
cbind(gama, alpha)



ajust3<-survreg(Surv(tempos,cens)~1,dist="lognorm")
ajust3

 require(survival)
 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)
 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)
 grupos<-c(rep(1,15),rep(2,14))
 ekm<- survfit(Surv(tempos,cens)~grupos)


 time<-ekm$time
 st<-ekm$surv
 ste<- exp(-time/20.41)
 stw<- exp(-(time/21.34)^1.54)
 stln<- pnorm((-log(time)+ 2.72)/0.76)
 stll<-1 / (1+(time/alpha)^lambda)

 cbind(time,st,ste,stw,stln, stll)


  par(mfrow=c(1,3))
  plot(ste,st,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), ylab = "S(t): Kaplan-Meier",
        xlab="S(t): exponencial")
```

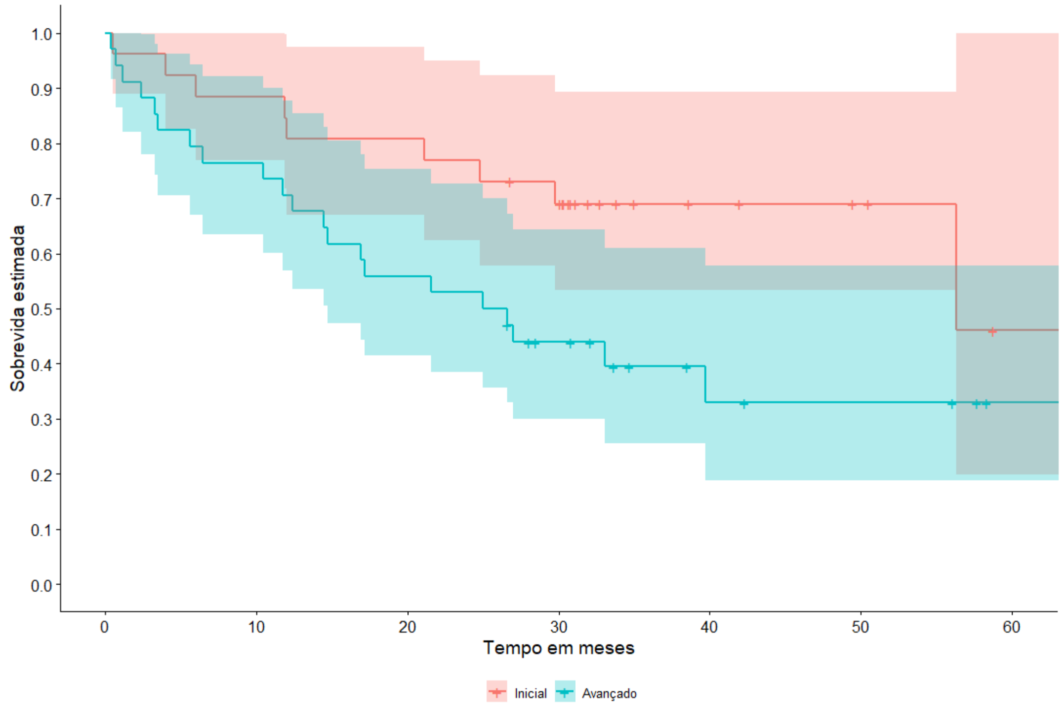

Figura 4: Estimativas por estágio da doença da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

```
  lines(c(0,1), c(0,1), type="l", lty=1)
  plot(stw,st,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), ylab = "S(t): Kaplan-Meier",
         xlab="S(t): Weibull")
  lines(c(0,1), c(0,1), type="l", lty=1)
  plot(stln,st,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), ylab = "S(t): Kaplan-Meier",
         xlab="S(t): log-normal")
  lines(c(0,1), c(0,1), type="l", lty=1)
```

7.2 Códigos utilizados no Cnário 2

```
require(survival)

tempos<-c(3,5,6,7,8,9,10,10,12,15,15,18,19,20,22,25,28,30,40,45)
cens<-c(1,1,1,1,1,1,1,0,1,1,0,1,1,1,1,1,1,1,1,0)

ajust1<-survreg(Surv(tempos,cens)~1,dist="exponential")
ajust1
alpha<-exp(ajust1$coefficients[1])
alpha

ajust2<-survreg(Surv(tempos,cens)~1,dist="weibull")
ajust2

alpha<-exp(ajust2$coefficients[1])

gama<-1/ajust2$scale
cbind(gama, alpha)



ajust3<-survreg(Surv(tempos,cens)~1,dist="lognorm")
ajust3

 require(survival)
 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)
 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)
 grupos<-c(rep(1,15),rep(2,14))
 ekm<- survfit(Surv(tempos,cens)~grupos)
```

```r


time<-ekm$time
st<-ekm$surv
ste<- exp(-time/20.41)
stw<- exp(-(time/21.34)^1.54)
stln<- pnorm((-log(time)+ 2.72)/0.76)
stll<-1 / (1+(time/alpha)^lambda)

cbind(time,st,ste,stw,stln, stll)


 par(mfrow=c(1,3))
 plot(ste,st,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), ylab = "S(t): Kaplan-Meier",
        xlab="S(t): exponencial")
 lines(c(0,1), c(0,1), type="l", lty=1)
 plot(stw,st,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), ylab = "S(t): Kaplan-Meier",
        xlab="S(t): Weibull")
 lines(c(0,1), c(0,1), type="l", lty=1)
 plot(stln,st,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), ylab = "S(t): Kaplan-Meier",
        xlab="S(t): log-normal")
 lines(c(0,1), c(0,1), type="l", lty=1)
```