UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
DEPARTAMENTO DE ESTATÍSTICA
ANÁLISE DE SOBREVIVÊNCIA

Abordagens Paramétricas e Não-Paramétricas em Análise de Sobrevivência

Aluna: Vitória Nascimento de Jesus Sesana

Professora: Agatha Sacramento Rodrigues

Vitória - ES
2024

Sumário

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 T, chamada como tempo de falha, uma variável aleatória que mede o tempo até a ocorrência de um evento, a função sobrevivência descreve a probabilidade da observação não falhar após o tempo t. O interesse principal é estimar a função de sobrevivência, a partir dela consegue-se obter outras médidas úteis para a interpretação do contexto do problema em estudo.

2.2 Estimador Kaplan-Meier

O estimador de Kaplan-Meier, ou estimador limite-produto, é um dos métodos não-paramétricos 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.

2.3 Teste de Log-Rank

É um teste não paramétrico utilizado na comparação de curvas de sobrevida entre dois ou mais grupos.

2.4 Modelos probabilísticos

Uma alternativa para estimar a função de sobrevivência é supor que a variável tempo de falha seja descrita como uma distribuição de probabilidade. A partir disso, utiliza-se métodos inferenciais, como o estimador de máxima verossimilhança, para estimar parâmetros da distribuição de probabilidade e posteriormente estimar a função de sobrevivência. O método da máxima verossimilhança é o mais adequado, pois ele considera censura em suas estimativas.

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.

Em 1997, Bartolucci e Fraser 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 óbito (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 da taxa de falha é *dead*, as demais são características adicionais dos pacientes, denominadas covariáveis.

Com o propósito de analisar o comportamento do tempo de sobrevida dos pacientes entre grupos, será calculado as estimativas pelo método de Kaplan-Meier nos níveis das covariáveis *sex*, *age*, *hist* e *stage*. 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”).

Tabela 1: Medidas descritivas sobre o tempo de vida, em meses, até o óbtio de pacientes com doença de Hodkings.

	Mínimo	Média	Mediana	Desvio Padrão	Máximo
Tempo	0.37	27.35	29.12	17.46	66.13

3.1 Kaplan-Meier por Gêneros

Na Tabela 2 é observado a relação entre quantidade de eventos que apresentaram a taxa de falha, óbito pela doença de Hodkings, e o gênero dos pacientes. Percebe-se que, entre os gêneros, há mais pacientes masculinos (70%) do que femininos (30%).

Tabela 2: Valores cruzados de quantidade e proporções entre a ocorrência de óbito e gênero.

	Obito não observado	Obito observado	Total
Feminino	9 (15.0%)	9 (15.0%)	18 (30.0%)
Masculino	21 (35.0%)	21 (35.0%)	42 (70.0%)
Total	30 (50.0%)	30 (50.0%)	60 (100.0%)

A Figura 1 apresenta as funções de sobrevivência estimada por gênero feminino e masculino. É observável que as funções de sobrevida se cruzam em diversos pontos ao longo do tempo.

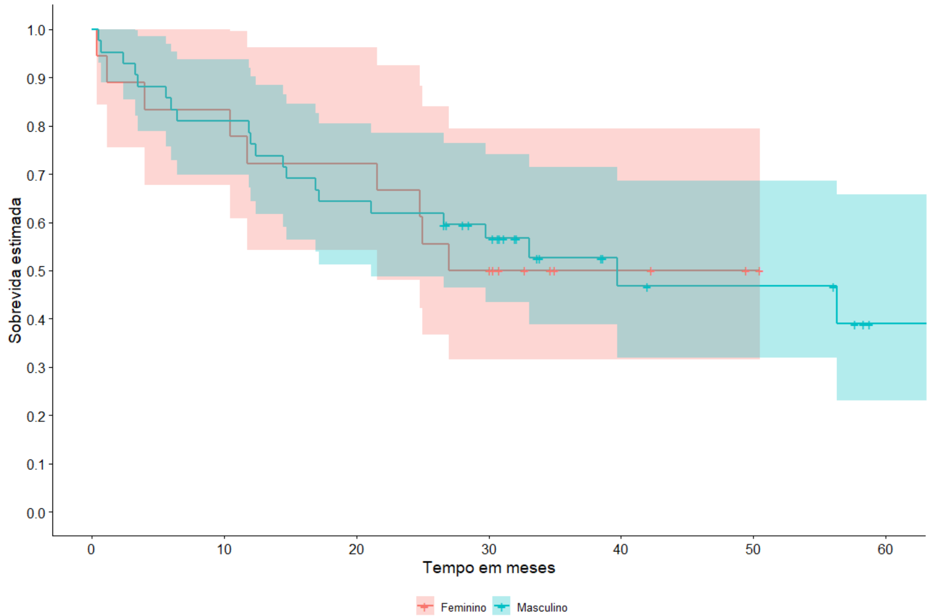


Figura 1: Estimativas por gênero da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

Entretanto, os resultados apresentados na Tabela 3 concluiram que não foram encontradas evidências entre diferenças nas funções de sobrevivência estimadas.

Tabela 3: Resultados teste de Log-Rank para comparações entre gêneros

Estatística de teste	Valor p
0.0303	0.8618

3.2 Kaplan-Meier por Faixas Etárias

A Tabela **??** apresenta iformações por nível de faixa etária. Percebe-se que a proporção entre os níveis não apresentam variação.

??

Tabela 4: Valores cruzados de quantidade e proporções entre a ocorrência de óbito e faixa etária.

	Obito não observado	Obito observado	Total
53 anos ou mais	7 (11.7%)	9 (15.0%)	16 (26.7%)
38 anos até menos 53 anos	9 (15.0%)	5 (8.3%)	14 (23.3%)
25 anos até menos 38 anos	4 (6.7%)	11 (18.3%)	15 (25.0%)
menos de 25 anos	10 (16.7%)	5 (8.3%)	15 (25.0%)
Total	30 (50.0%)	30 (50.0%)	60 (100.0%)

A Figura 2 apresenta as estimativas da função de sobrevida por Kaplan-Meier por nível de faixa etária. Percebe-se que para os pacientes de 25 até 38 anos apresentam decaimento da curva mais acelerado que dos demais níveis, indicando que pacientes nessa faixa etária tendem a apresentar a taxa de falha em menor tempo.

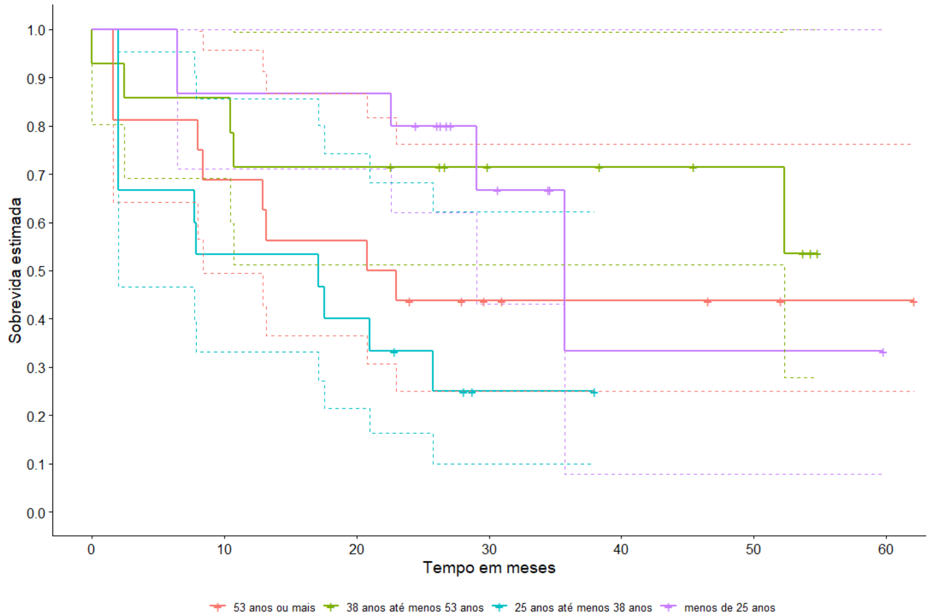


Figura 2: Estimativas por faixa etária da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

Como mostra a Tabela 6, os resultados indicam que há diferenças significativas entre os níveis da faixa etária.

Tabela 5: Resultados teste de Log-Rank para comparações entre faixas etárias.

Estatística de teste	Valor p
9.2433	0.0262

3.3 Kaplan-Meier por Tipos de Linfomas

Percebe-se, na Tabela 6, que há poucos pacientes com o fepleção de linfócitos (11.7%) e mais pacientes com esclerose nodular (53,3%).

Tabela 6: Valores cruzados de quantidade e proporções entre a ocorrência de óbito e tipo de linfoma.

	Óbito não observado	Óbito observado	Total
Esclerose nodular	18 (30.0%)	14 (23.3%)	32 (53.3%)
Fepleção de linfócitos	1 (1.7%)	6 (10.0%)	7 (11.7%)
Misto celular	11 (18.3%)	10 (16.7%)	21 (35.0%)
Total	30 (50.0%)	30 (50.0%)	60 (100.0%)

Na Figura 3, percebe-se que o decaimento da curva de sobrevivência estimada para os indíviduos com fepleção de linfócitos é mais acelerada do que para os demais tipos de linfócitos, ou seja, índividuos com este tipo de linfoma apresentaram o tempo de falha em menor tempo do que os demais, o que indicaria que as curvas de sobrevida são diferentes entre si.

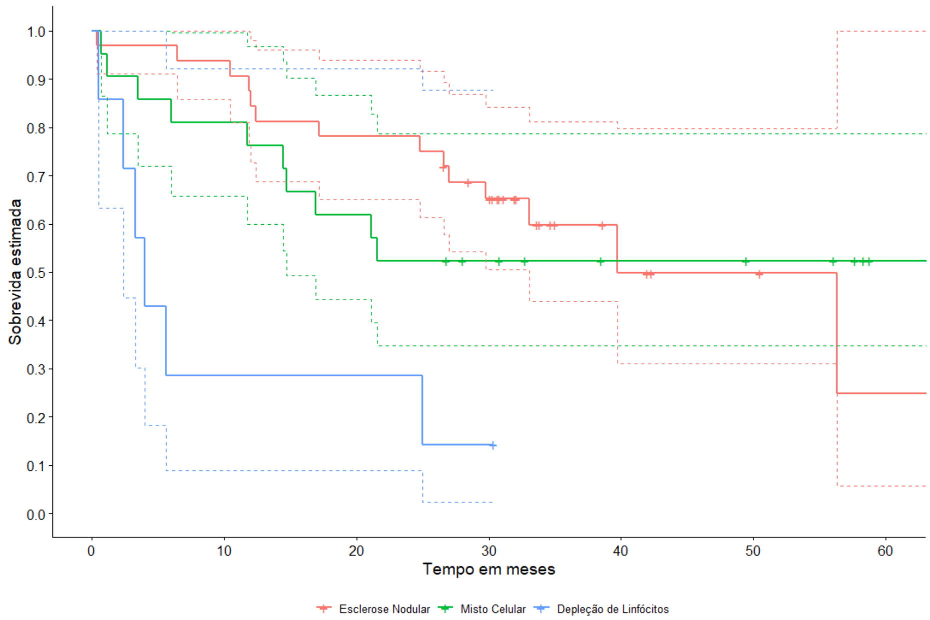


Figura 3: Estimativas por tipo de linfoma da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

A Tabela 8, assim como na Tabela 6, indica que há diferença nas curvas de sobrevida estimada entre os tipos de linfoma.

Tabela 7: Resultados teste de Log-Rank para comparações entre tipos de linfomas.

Estatística de teste	Valor p
12.7973	0.0016

3.4 Kaplan-Meier por Estágios da Doença

Observando pela perspectiva dos estágio da doença, há mais pacientes no estágio mais avançado do que (56.7%) em estágios iniciais (43.3%).

Tabela 8: Valores cruzados de quantidade e proporções entre a ocorrência de óbito e estágio da doença.

	Óbito não observado	Óbito observado	Total
Avançado	13 (21.7%)	21 (35.0%)	34 (56.7%)
Inicial	17 (28.3%)	9 (15.0%)	26 (43.3%)
Total	30 (50.0%)	30 (50.0%)	60 (100.0%)

Observando a Figura 4, verifica-se que as estimativas da função de sobrevida para pacientes com o estágio avançado da doença apresentam decaimento maior que a de pacientes em estágio inicial, indicando que as curvas de sobrevida entre os estágios sejam diferentes.

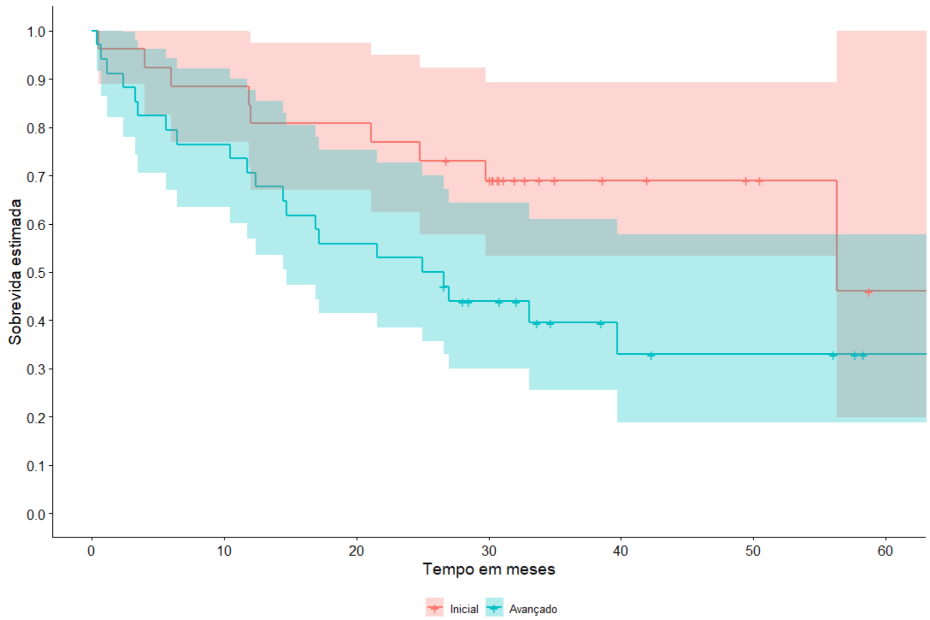


Figura 4: Estimativas por estágio da doença da função de sobrevivência de pacientes com a doença de Hodkings utilizando Kaplan-Meier.

Na Tabela 9, há evidências que indicam a diferença entre as curvas de sobrevivência estimadas entre o estágio inicial da doença e o estágio final.

Tabela 9: Resultados teste de Log-Rank para comparações entre estágios da doença.

Estatística de teste	Valor p
4.1425	0.0418

4 Cenário 2: Resistência de Isolentes Térmicos

Realizou-se um experimento para verificar o tempo de resitência de isolantes elétricos sob alta tensão. A parti disso, 25 unidades de isolantes elétricos foram submetidos a uma tensão de 35Kvolts até que 15 deles falhassem. O tempo até a ruptura dos 15 isolantes elétricos são motrados na Tabela 10.

Tabela 10: Tempo da ruptura de 15 isolantes elétricos, em minutos, submetidos a uma tensão de 35 Kvolts.

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	

Neste experimento, o tipo de censura apresentado é censura à direita tipo II com 15 observações que apresentassem a ruptura durante o experimento.

4.1 Modelos Paramétricos

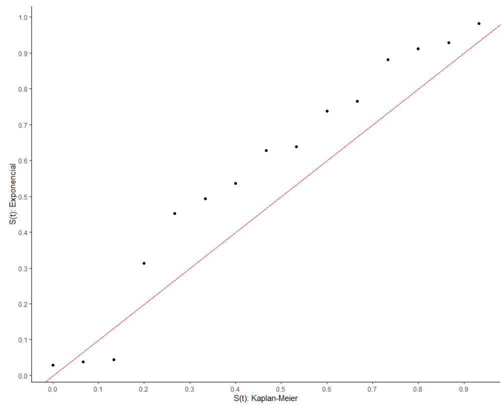
Para realizar suposições a respeito do tempo de ruptura dos isolantes elétricos utilizados no problema, foram estimadas funções de sobrevivências nos tempos coletados para quatro modelos probabilísticos a fim de selecionar aquele que mais se adequa aos dados. As estimativas são apresentados na Tabela 11.

Tabela 11: Estimativas de sobrevivência para os tempos de ruptura de isolantes térmicos usando estimadores de Kaplan-Meier e os modelos exponencial, Weibull, log-normal e log-logística.

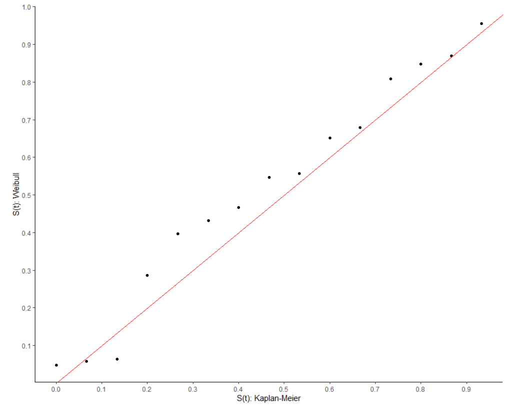
	Tempo	Kaplan-Meier	Exponencial	Weibull	Log-Normal	Log-Logística
1	0.19	0.93	0.98	0.96	0.99	0.98
2	0.78	0.87	0.93	0.87	0.89	0.90
3	0.96	0.80	0.91	0.85	0.86	0.87
4	1.31	0.73	0.88	0.81	0.80	0.83
5	2.78	0.67	0.77	0.68	0.63	0.66
6	3.16	0.60	0.74	0.65	0.60	0.62
7	4.67	0.53	0.64	0.56	0.49	0.51
8	4.85	0.47	0.63	0.55	0.48	0.50
9	6.50	0.40	0.54	0.47	0.40	0.41
10	7.35	0.33	0.49	0.43	0.37	0.38
11	8.27	0.27	0.45	0.40	0.34	0.34
12	12.07	0.20	0.31	0.29	0.25	0.25
13	32.52	0.13	0.04	0.06	0.09	0.09
14	33.91	0.07	0.04	0.06	0.08	0.09
15	36.71	0.00	0.03	0.05	0.07	0.08

O estimador de Kaplan-Meier também foi calculado para servir de comparação com os modelos, aquele que mais se assemelha as estimativas de Kaplan-Meier será o mais adequado.

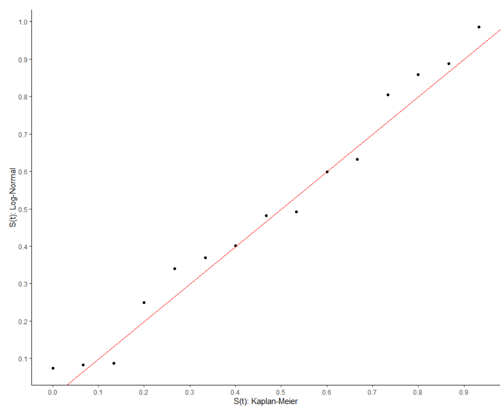
Uma das alternativas para verifica é comparar gráficamente a função de sobrevida estimada do Kaplan-Meier com a função de sobrevida estimada pelo modelo. Diante disso, os resultados são apresentados na Figura .



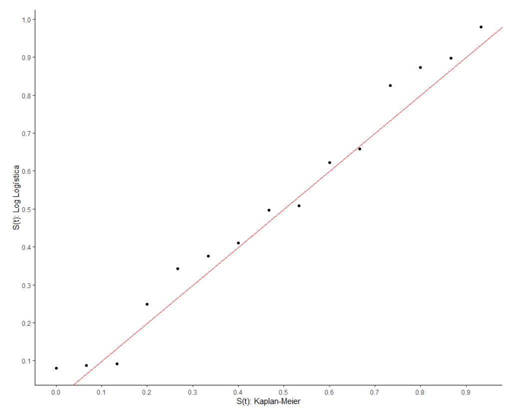
(a) Modelo exponencial.



(b) Modelo Weibull.



(c) Mdelo log-normal.



(d) Modelo log-logística.

Figura 5: Gráficos das sobrevivências estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modeloes exponencial, Weibull, log-normal e log-logística

A Figura 5 indica que os modelos log-normal, Subfigura 5c, e log-logistica, Subfigura 5d, apresentaram melhor desempenho em se ajustar as estimativas por Kaplan-Meier.

Para critério de desempate, a seleção do modelo será realizada com base no menor erro quadrático médio. Desse modo, observando os resultados da Tabela 12, o modelo log-normal (0.00212) é o modelo que melhor representa o comportamento da variável de interesse do problema apresentado.

Tabela 12: Erro quadrático médio dos modelos exponencial, Weibull, log-normal e log-logística em relação as estimativas de Kaplan-Meier.

	Exponencial	Weibull	Log-Normal	Log-Logística
EQM	0.01379	0.00425	0.00212	0.00251

As estimativas dos parâmetros do modelo log-normal, em escala original, e seus intervalos são mostrados na Tabela 13. Obteve-se que a média, um dos parâmetros da distribuição log-normal, é de 1.51, IC[μ; 95%] = (0.78,2.25), e o desvio padrão, outro parâmetro da distribuição log-normal, estimado em 1.45, IC[σ; 95%] = (1.01, 2.07)

Tabela 13: Estimativa e intervalo de confiança da média e desvio padrão do modelo log-normal em escala original.

	Estimativa	L95%	U95%
Média	1.51	0.78	2.25
Desvio padrão	1.45	1.01	2.07

Com o modelo selecionado e seus parâmetros estimados, a função de sobrevivência estimada no tempo t é dada por:

$$
\begin{aligned}
S(t) &= \phi\left(\frac{-log(t)+\mu}{\sigma}\right) \\
\widehat{S(t)} &= \phi\left(\frac{-log(t)+1.51}{1.45}\right)
\end{aligned} \tag{1}
$$

4.1.1 Tempo Mediano

O percentil da distribuição log-normal podem ser obtidos utilizando a tabela normal padrão com base na seguinte equação:

$$
t_p = exp\{z_p\sigma + \mu\} \tag{2}
$$

Para o tempo mediano, o cálculo se dar por:

$$
\begin{aligned}
\widehat{t_{0.5}} &= exp\{z_{0.5}1.45 + 1.51\} \\
&= 4.54
\end{aligned} \tag{3}
$$

4.1.2 Tempo Médio

A estimativa do tempo médio de vida deste tipo de isolante elétrico funcionando a 35 Kvolts para o modelo selecionado é dado por:

$$
E[T] = exp\{\mu + \sigma^2/2\} \tag{4}
$$

Desse modo, o tempo de vida médio estimado é de:

$$
\begin{aligned}
\widehat{E[T]} &= exp\{1.51 + (1.45)^2/2\} \\
&= 12.96
\end{aligned} \tag{5}
$$

4.1.3 Tempo de ruptura

O tempo necessário para que 20% dos isolantes estarem fora de operação é a função de distribuição da log normal no tempo 20, ou seja, $F_{ln}(t) = 0.20$.

$$
\begin{aligned}
F_{ln}(t) &= \phi\left(\frac{log(t)-\mu}{\sigma}\right) \\
0.2 &= F_z\left(\frac{log(t)-\mu}{\sigma}\right) \\
0.2 &= P\left(Z \leq \frac{log(t)-\mu}{\sigma}\right) \\
0.2 &= P\left(Z \leq z_{0.2}\right)
\end{aligned} \tag{6}
$$

Sendo $z_{0.2}$ = -0.8416212.

$$
\begin{aligned}
z_{(0.2)} &= \frac{log(t)-\mu}{\sigma} \\
t &= e\left(z_{(0.2)}\sigma + \mu\right) \\
\widehat{t} &= e(-0.8416212 * 1.45 + 1.51) \\
\widehat{t} &= 1.341276
\end{aligned} \tag{7}
$$

5 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.

6 Apêndices

6.1 Códigos utilizados no Cenário 1

```
# AN LISE DE SOBREVIV NCIA
# EXERCiCIO 4 DA LISTA 2

# bibliotecas -------------------------------------------------------------
require(survival)
require(survminer)
require(summarytools)
require(dplyr)

# base de dados -----------------------------------------------------------
base <- readxl::read_xlsx("AS_lista2/Lista2_Hodgkins.xlsx") %>%
  mutate(aux_age = case_when(
    (age > 53) ~ "53 anos ou mais",
    (age >= 38 & age < 53) ~ "38 anos at   menos 53 anos",
    (age >= 25 & age < 38) ~ "25 anos at   menos 38 anos",
    (age < 25) ~ "menos de 25 anos",
  ),
  aux_dead = case_when(
    dead == 0 ~ "Evento n o observado",
    dead == 1 ~ "Evento observado"
  ),
  aux_sex = case_when(
    sex == 0 ~ "Feminino",
    sex == 1 ~ "Masculino"
  ),
  aux_stage = case_when(
    stage == 0 ~ "Inicial",
    stage == 1 ~ "Avan ado"
  ),
  aux_hist = case_when(
    hist == 1 ~ "Esclerose nodular",
    hist == 2 ~ "Misto celular",
    hist == 3 ~ "Feple   o de linf citos"
  ),
  aux_age = factor(aux_age, labels = c("53 anos ou mais",
                       "38 anos at   menos 53 anos",
                       "25 anos at   menos 38 anos",
                       "menos de 25 anos"))
  )

# analise descritiva ------------------------------------------------------
ctable(base$aux_sex, y=base$aux_dead, prop="t")
ctable(base$aux_age, y=base$aux_dead, prop="t")
ctable(base$aux_hist, y=base$aux_dead, prop="t")
ctable(base$aux_stage, y=base$aux_dead, prop="t")

ggplot(base) +
  aes(x = survivaltime, y = aux_dead) +
  geom_boxplot(fill = "#0c4c8a") +
  labs(x = "Tempo de vida dos pacientes", y = "G nero") +
  theme_minimal()

# descritiva survivaltime -------------------------------------------------

min(base$survivaltime) # minimo em meses
```

```r
min(base$survivaltime)*30 # minimo em dias
max(base$survivaltime) # m ximo em meses
max(base$survivaltime)/12 #m ximo em dias

boxplot(base$survivaltime, horizontal = TRUE)

descr(base$survivaltime, stats = c("min", "mean", "med","sd","max"),
      transpose = TRUE)

# tabela observa  es e censuras -----------------------------------------
Surv(time = base$survivaltime, event = base$dead)

# genero --------------------------------------------------------------
ajuste_sexo <-
  survfit(
    Surv(time = survivaltime,
         event = dead) ~ aux_sex,
    data = base
    )

ajuste_sexo <- survfit(Surv(time = survivaltime,
                            event = dead) ~ sex,
                       data = base)

#Para fazer a tabela da estimativa de KM e intervalo de confian a:
surv_summary(ajuste_sexo, data = lung)

# estimativas sobrevida por kaplan-meier
kp_sexo <- surv_summary(ajuste_sexo,
                        data = base)

# gr fico sobreviva estimada
gg_sexo <- survminer::ggsurvplot(ajuste_sexo, data = base,
                      pval = FALSE, conf.int=TRUE, conf.int.style = "ribbon",
                      ylab = "Sobrevida estimada",
                      xlab = "Tempo em meses",
                      legend.title = "",
                      legend.labs = c("Feminino","Masculino"))

gg_sexo$plot <- gg_sexo$plot +
  scale_y_continuous(limits = c(0, 1),
                     breaks = seq(0, 1, by = 0.1)) +
  scale_x_continuous(limits = c(0, max(base$survivaltime)),
                     breaks = seq(0, max(base$survivaltime), by = 10)) +
  theme(legend.position = "bottom")

# teste log-rank
test_sexo <-survdiff(Surv(time = survivaltime,
                          event = dead) ~ sex,
                     data = base)

# faixa etaria ---------------------------------------------------------
ajuste_age <- survfit(
  Surv(time = survivaltime,
       event = dead,
       type = "right",
       origin = 4) ~ aux_age,
  data = base)

# estimativas sobrevida por kaplan-meier
kp_age <- surv_summary(ajuste_age, data = base)

# gr fico sobreviva estimada
gg_age <- survminer::ggsurvplot(ajuste_age, data = base,
                                 pval = FALSE, conf.int=TRUE,
                                conf.int.style = "step",
                                 ylab = "Sobrevida estimada",
                                 xlab = "Tempo em meses",
                                 legend.title = "",
                                 legend.labs =  c("53 anos ou mais",
                                                  "38 anos at   menos 53 anos",
                                                  "25 anos at   menos 38 anos",
                                                  "menos de 25 anos"))
```

```
gg_age$plot <- gg_age$plot +
  ggplot2::scale_y_continuous(limits = c(0, 1),
                              breaks = seq(0, 1, by = 0.1)) +
  ggplot2::scale_x_continuous(limits = c(0, max(base$survivaltime)),
                              breaks = seq(0, max(base$survivaltime), by = 10)) +
  theme(legend.position = "bottom") +
  scale_fill_discrete(labels=c('High Program', 'Low Program', 'c', 'd'))


# teste log-rank
test_age <- survdiff(Surv(time = survivaltime,
                            event = dead) ~ aux_age,
                       data = base)


# histologia ---------------------------------------------------------------
ajuste_hist <- survfit(Surv(time = survivaltime,
                            event = dead) ~ hist,
                       data = base)

# estimativas sobrevida por kaplan-meier
kp_hist <- surv_summary(ajuste_hist, data = base)

# gr fico sobreviva estimada
gg_hist <- survminer::ggsurvplot(ajuste_hist,
                                 data = base,
                                 pval = FALSE,
                                 conf.int=TRUE,
                                 conf.int.style = "step",
                                 ylab = "Sobrevida estimada",
                                 xlab = "Tempo em meses",
                                 legend.title = "",
                                 legend.labs = c("Esclerose Nodular",
                                                 "Misto Celular",
                                                 "Deple  o de Linf citos"))

gg_hist$plot

gg_hist$plot <- gg_hist$plot +
  scale_y_continuous(limits = c(0, 1),
                     breaks = seq(0, 1, by = 0.1)) +
  scale_x_continuous(limits = c(0, max(base$survivaltime)),
                     breaks = seq(0, max(base$survivaltime), by = 10)) +
  theme(legend.position = "bottom")

# teste log-rank
test_hist <- survdiff(Surv(time = survivaltime,
                           event = dead) ~ hist,
                      data = base)

# estagio da doen a -------------------------------------------------------
ajuste_stage <- survfit(Surv(time = survivaltime,
                             event = dead) ~ stage,
                        data = base)

# estimativas sobrevida por kaplan-meier
kp_stage <- surv_summary(ajuste_stage, data = base)

# gr fico sobreviva estimada
gg_stage <- survminer::ggsurvplot(ajuste_stage,
                                  data = base,
                                  pval = FALSE,
                                  conf.int=TRUE,
                                  conf.int.style = "ribbon",
                                  ylab = "Sobrevida estimada",
                                  xlab = "Tempo em meses",
                                  legend.title = "",
                                  legend.labs = c("Inicial",
                                                  "Avan ado"))

gg_stage$plot <- gg_stage$plot +
  scale_y_continuous(limits = c(0, 1),
```

```
                     breaks = seq(0, 1, by = 0.1)) +
  scale_x_continuous(limits = c(0, max(base$survivaltime)),
                     breaks = seq(0, max(base$survivaltime), by = 10)) +
  theme(legend.position = "bottom")

# teste log-rank
test_stage <- survdiff(Surv(time = survivaltime,
                            event = dead) ~ stage,
                       data = base)
```

6.2 Códigos utilizados no Cenário 2

```
# ANaLISE DE SOBREVIVeNCIA
# EXERCiCIO 4 DA LISTA 3

# bibliotecas -----------------------------------------------------------
require(survival)
require(survminer)
require(summarytools)
require(dplyr)
require(flexsurv)

# base de dados ---------------------------------------------------------
base <-
  cbind(
    c(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),
    rep(1, 15))

colnames(base) <- c("temp_falha", "ind_falha")
base <- as_tibble(base)

# kaplan meier ----------------------------------------------------------
KaplanMeir <-
  survfit(
    Surv(time = temp_falha,
         event = ind_falha) ~ 1,
    data = base
    )

st_KaplanMeier <- KaplanMeir$surv

# exponential -----------------------------------------------------------
modelExponential <-
  flexsurvreg(
    Surv(
      time = temp_falha,
      event = ind_falha
    ) ~ 1,
    data = base,
    dist = "exponential"
  )

st_Exponencial <- as.data.frame(summary(modelExponential))$est

# weibull ---------------------------------------------------------------
modelWeibull <-
  flexsurvreg(
    Surv(
      time = temp_falha,
      event = ind_falha
    ) ~ 1,
    data = base,
    dist = "weibull"
  )

st_Weibull <- as.data.frame(summary(modelWeibull))$est

# log normal ------------------------------------------------------------
modelLogNorm <-
  flexsurvreg(
    Surv(
      time = temp_falha,
```

```
      event = ind_falha
      ) ~ 1,
    data = base,
    dist = "lognorm"
  )

st_LogNorm <- as.data.frame(summary(modelLogNorm))$est

# log logistic ------------------------------------------------------------
modelLogLogistic <-
  flexsurvreg(
    Surv(
      time = temp_falha,
      event = ind_falha
    ) ~ 1,
    data = base,
    dist = "llogis"
  )

st_LogLogistic <- as.data.frame(summary(modelLogLogistic))$est

# gr ficos comparativos ---------------------------------------------------

# base com as estimativas de KP, EXP, WEIBULL, LOGNORM e LOGLOGISTICA
sts <- cbind(
  base$temp_falha,
  st_KaplanMeier,
  st_Exponencial,
  st_Weibull,
  st_LogNorm,
  st_LogLogistic
  ) %>%
  as_tibble()

# KP x EXP
ggExp <- sts %>%
ggplot(aes(x = st_KaplanMeier,
           y = st_Exponencial)) +
  geom_point() +
  xlab("S(t): Kaplan-Meier") +
  ylab("S(t): Exponencial") +
  theme_classic() +
  geom_abline(color = "red") +
  scale_x_continuous(breaks = seq(0,1, 0.1)) +
  scale_y_continuous(breaks = seq(0,1, 0.1))

# KP x WEIBULL
ggWeib <- sts %>%
  ggplot(aes(x = st_KaplanMeier,
             y = st_Weibull)) +
  geom_point() +
  xlab("S(t): Kaplan-Meier") +
  ylab("S(t): Weibull") +
  theme_classic() +
  geom_abline(color = "red") +
  scale_x_continuous(breaks = seq(0,1, 0.1)) +
  scale_y_continuous(breaks = seq(0,1, 0.1))

# KP x LN
ggLN <- sts %>%
  ggplot(aes(x = st_KaplanMeier,
             y = st_LogNorm)) +
  geom_point() +
  xlab("S(t): Kaplan-Meier") +
  ylab("S(t): Log-Normal") +
  theme_classic() +
  geom_abline(color = "red") +
  scale_x_continuous(breaks = seq(0,1, 0.1)) +
  scale_y_continuous(breaks = seq(0,1, 0.1))

# KP x LL
ggLL <- sts %>%
  ggplot(aes(x = st_KaplanMeier,
```

```
            y = st_LogLogistic)) +
  geom_point() +
  xlab("S(t): Kaplan-Meier") +
  ylab("S(t): Log Log stica") +
  theme_classic() +
  geom_abline(color = "red") +
  scale_x_continuous(breaks = seq(0,1, 0.1)) +
  scale_y_continuous(breaks = seq(0,1, 0.1))

# erro quadr tico
eq <- sts %>%
  mutate(eq_Exp2 = (st_KaplanMeier - st_Exponencial)^2,
         eq_Wei2 = (st_KaplanMeier - st_Weibull )^2,
         eq_LN2 = (st_KaplanMeier - st_LogNorm )^2,
         eq_LL2 = (st_KaplanMeier - st_LogLogistic)^2,
         ) %>%
  select(eq_Exp2, eq_Wei2, eq_LN2, eq_LL2)

# erro quadr tico m dio
mqe <- apply(eq, 2, sum)/length(base$temp_falha)

# tabela par metros estimados
est <- modelLogNorm$res # escala original
est_log <-modelLogNorm$res.t # escala log

muln <- modelLogNorm$res[1,1]
sdln <- modelLogNorm$res[2,1]

# tempo mediano
t_0.5 = exp( (qnorm(0.5) * sdln) + muln)

# tempo medio
tm = exp( muln + ( (sdln^2) / 2 ) )

# tempo isolante 20% rupturados
z_0.2 <- qnorm(0.2)
t = exp(muln + (sdln * z_0.2) )
```