



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE ESTATÍSTICA**

**PAMELLA KARDYNNALE DE OLIVEIRA SILVA**

**ESTUDO DE RESÍDUOS NA ANÁLISE DE SOBREVIVÊNCIA COM  
APLICAÇÃO EM DADOS DE TRATAMENTO DO TIPO BIOFEEDBACK**

**CAMPINA GRANDE-PB**

**2021**

PAMELLA KARDYNNALE DE OLIVEIRA SILVA

**ESTUDO DE RESÍDUOS NA ANÁLISE DE SOBREVIVÊNCIA COM  
APLICAÇÃO EM DADOS DE TRATAMENTO DO TIPO BIOFEEDBACK**

Trabalho de Conclusão de Curso (Artigo) apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

**Orientador:** Tiago Almeida de Oliveira  
**Coorientador:** Pollyanna Kelly de Oliveira Silva

**CAMPINA GRANDE - PB  
2021**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586e Silva, Pamella Kardynnale de Oliveira.  
Estudo de resíduos na análise de sobrevivência com aplicação em dados de tratamento do tipo *biofeedback* [manuscrito] / Pamella Kardynnale de Oliveira Silva. - 2021.  
35 p.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2021.

"Orientação : Prof. Dr. Tiago Almeida de Oliveira, Coordenação do Curso de Estatística - CCT."

1. Análise de sobrevivência. 2. Estimador de Kaplan-Meier.  
3. Modelo de Cox. I. Título

21. ed. CDD 519.536

PAMELLA KARDYNNALE DE OLIVEIRA SILVA

ESTUDO DE RESÍDUOS NA ANÁLISE DE SOBREVIVÊNCIA COM  
APLICAÇÃO EM DADOS DE TRATAMENTO DO TIPO BIOFEEDBACK

Trabalho de Conclusão de Curso  
(Artigo) apresentado ao curso de  
Bacharelado em Estatística do  
Departamento de Estatística do  
Centro de Ciências e Tecnologia da  
Universidade Estadual da Paraíba  
em cumprimento às exigências legais  
para obtenção do título de bacharel  
em Estatística.

Aprovada em: 06/08/2021

**BANCA EXAMINADORA**



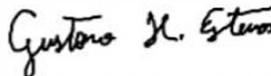
---

Prof. Dr. Tiago Almeida de Oliveira (Orientador)  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Me. Ednário Barbosa de Mendonça  
Universidade Estadual da Paraíba (UEPB)



---

Prof. Dr. Gustavo Henrique Esteves  
Universidade Estadual da Paraíba (UEPB)

Primeiramente agradeço a Deus por mais essa realização e a minha família que com muito apoio não mediram esforços para que eu chegasse até essa etapa da vida.

“Existem dois tipos de estatísticas: as  
que você lê e as que você faz”.  
(Rex St)

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> – Curva de Kaplan Meier para os tratamentos.....	<b>25</b>
<b>Figura 2</b> - Representação gráfica das estimativas de Kaplan-Meier acumulada para a função de sobrevivência para os tratamentos.....	<b>25</b>
<b>Figura 3</b> - Gráfico do uso dos resíduos padronizados de Schoenfeld.....	<b>28</b>
<b>Figura 4</b> - Gráfico dos resíduos Martingale e Deviance.....	<b>29</b>
<b>Figura 5</b> - Gráfico do resíduo escore.....	<b>30</b>

## LISTA DE TABELAS

<b>Tabela 1</b> – Estimativas do Kaplan-Meier para o tempo de recuperação em paciente com o tratamento Conservador (bfb=0) .....	23
<b>Tabela 2</b> - Estimativas do Kaplan-Meier para o tempo de recuperação em paciente com o tratamento <i>Biofeedback</i> (bfb=1) .....	24
<b>Tabela 3</b> - Teste de Log-Rank para os tratamentos em estudo.....	26
<b>Tabela 4</b> - Teste de proporcionalidade das funções de risco no modelo ajustado .....	27
<b>Tabela 5</b> - Estimativa do parâmetro referente ao modelo estratificado.....	27
<b>Tabela 6</b> - Teste de proporcionalidade da função de risco do modelo estratificado.....	27



## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	09
<b>2 REVISÃO DE LITERATURA</b> .....	10
<b>2.1 Análise de sobrevivência</b> .....	10
<b>2.1.1 Censura</b> .....	11
2.1.1.1 Censura à Direita.....	11
2.1.1.1.1 Censura do tipo I.....	12
2.1.1.1.2 Censura do tipo II.....	12
2.1.1.1.3 Censura Aleatória.....	12
2.1.1.2 Censura à Esquerda.....	12
2.1.1.3 Censura Intervalar.....	13
<b>2.2 Função de sobrevivência</b> .....	13
<b>2.2.1 Função densidade de probabilidade (f(t))</b> .....	14
<b>2.3 Função de risco</b> .....	14
<b>2.4 Estimador de Kaplan Meier</b> .....	16
<b>2.5 Teste de Log-Rank</b> .....	17
<b>2.6 Modelo de Cox</b> .....	18
<b>2.7 Análise de Resíduos</b> .....	19
2.7.1 Resíduos padronizados de Schoenfeld.....	19
2.7.2 Resíduos Martingale.....	20
2.7.3 Resíduos Deviance .....	21
2.7.4 Resíduos Escore .....	21
<b>3 MATERIAL E METÓDOS</b> .....	22
3.1 Banco de dados .....	22
<b>4 RESULTADOS</b> .....	22
<b>5 DISCUSSÃO</b> .....	31
<b>6 CONCLUSÃO</b> .....	32
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	33

## ESTUDO DE RESÍDUOS NA ANÁLISE DE SOBREVIVÊNCIA COM APLICAÇÃO EM DADOS DE TRATAMENTO DO TIPO BIOFEEDBACK

### RESIDUE STUDY IN SURVIVAL ANALYSIS WITH APPLICATION TO BIOFEEDBACK TYPE TREATMENT DATA

Pamella Kardynnale de O. S. <sup>1\*</sup>

Pollyanna Kelly de O. S. <sup>2</sup>

Tiago Almeida O. <sup>3\*</sup>

#### RESUMO

O trabalho apresentado tem como objetivo principal analisar o tempo de início do tratamento até o paciente atingir a reabilitação completa da deglutição, foi utilizado o modelo de regressão de Cox com o auxílio da ferramenta computacional R, aplicado a um banco de dados "*Biofeedback*" constituído de um total de 33 observações por um período de 100 dias. Neste trabalho, aplicou-se os métodos não-paramétricos de Kaplan-Meier e o teste de Log-Rank como também foi feita uma análise dos resíduos para verificar os pressupostos de proporcionalidade ao longo do tempo. Os resultados obtidos pelo método *stepwise* mostraram que as covariáveis tratamento e tempo transformado apresentaram melhor modelo e foi feito o teste de proporcionalidade que resultou no valor p menor ao nível de 5% de significância, concluindo assim que há evidências de violação de proporcionalidade. Para isso, foi utilizado o modelo de Cox estratificado para a variável tratamento e em seguida, o teste de proporcionalidade tendo como resultado o valor p próximo ao nível de 5% de significância, sendo assim o modelo de Cox estratificado é o mais indicado para uma análise gráfica de resíduos.

**Palavras-Chaves:** Análise de sobrevivência, Kaplan-Meier, Modelo de Cox.

---

<sup>1</sup>Bacharelada em Estatística do CCT-UEPB. Email: [pamellakardinnale@gmail.com](mailto:pamellakardinnale@gmail.com)

<sup>2</sup> Professora do DE-CCT-UEPB. Email: [pollyanna\\_kelly@yahoo.com.br](mailto:pollyanna_kelly@yahoo.com.br)

Professor do DE-CCT-UEPB. Email: [tadolive@servidor.uepb.edu.br](mailto:tadolive@servidor.uepb.edu.br)

## ABSTRACT

The main objective of the present work is to analyze the time from the beginning of the treatment until the patient reaches the complete rehabilitation of swallowing. The Cox regression model was used with the aid of the computational tool R, applied to a database "Biofeedback" constituted of a total of 33 observations over a period of 100 days. In this work, the non-parametric Kaplan-Meier methods and the Log-Rank test were applied, as well as an analysis of the residuals to verify the assumptions of proportionality over time. The results obtained by the stepwise method showed that the treatment and transformed time covariates presented a better model and the proportionality test was performed, which resulted in the p-value lower at the 5% level of significance, thus concluding that there is evidence of violation of proportionality. For this, the stratified Cox model was used for the treatment variable, followed by the proportionality test, resulting in a p-value close to the 5% level of significance, thus the stratified Cox model is the most suitable for a graphical residue analysis.

**Keywords:** Survival analysis, Kaplan-Meier, Cox model.

## 1 INTRODUÇÃO

A análise de sobrevivência é a área da estatística que mais evoluiu nas últimas décadas e tem como objetivo principal estudar o tempo de vida dos indivíduos desde o começo até que ocorra o evento de interesse, por exemplo, o tempo em que um indivíduo desenvolveu uma doença (início, meio e fim), a recuperação do indivíduo a um determinado tratamento, etc.

É comum o surgimento de desistência de indivíduos no tratamento, a não ocorrência do evento de interesse, tal como o tratamento do indivíduo foi suspenso antes do seu estudo final. Esse tipo de acontecimento é chamado de censura que são observações incompletas, isto é, quando por algum motivo não observamos o evento de interesse e mesmo sendo censurados fornecem informações sobre o tempo de vida dos indivíduos, e são classificadas em 3 tipos: Censura à direita, Censura à esquerda e Censura intervalar.

Nesse sentido, foi utilizado o estimador de Kaplan-Meier que tem como objetivo descrever e comparar o comportamento da função de sobrevivência para diferentes grupos, o teste de Log-Rank com o interesse de comparar funções de sobrevivência e o método do modelo Cox proposto em 1972 pelo estatístico britânico David Cox com a finalidade de verificar a influência das covariáveis no tempo até a ocorrência do evento de interesse. Diante disso, foi utilizado o banco de dados “*biofeedback*” com o objetivo de analisar os pacientes que sofrem de aspiração após uma cirurgia de cabeça e pescoço e para isso foi utilizado dois tipos de tratamentos para verificar qual o tempo de o início do tratamento até o paciente atingir a reabilitação completa da deglutição. O tratamento é iniciado assim que terminar o processo de cicatrização após a cirurgia, sendo assim, acompanhados durante 100 dias.

Apesar de terem sido propostos há muito tempo, as curvas de Kaplan-Meier e o modelo de Cox têm se tornado cada vez mais populares em pesquisas da área médica e, muitas vezes, são utilizados como a única alternativa possível para analisar dados dessa natureza. Embora eles expliquem satisfatoriamente muitas questões ou hipóteses clínicas, é comum encontrar problemas com algumas particularidades que limitam a aplicação desses métodos (RIGÃO et al., 2020).

Este trabalho tem como foco principal estudar as técnicas de análise de sobrevivência, como também demonstrar a importância da análise de resíduos e dos pressupostos do modelo de Cox e verificar qual o método é o mais indicado para esse estudo.

## **2 REVISÃO DE LITERATURA**

### **2.1 Análise de Sobrevivência**

Segundo Borges (2014) a análise de sobrevivência consiste, entre outras coisas, em analisar os tempos de vida dos indivíduos desde o seu momento de entrada no estudo, até o momento em que ocorre o acontecimento de interesse, acontecimento esse que é definido à priori. Este acontecimento é geralmente definido como uma falha, que pode ser morte, recaída de uma doença ou até mesmo quando um tratamento começa a fazer efeito no paciente. Tem a particularidade de lidar com dados censurados, ou

seja, em alguns dos indivíduos pode não chegar a ocorrer o acontecimento de interesse durante o período de observação. Indivíduos que continuam vivos após o término do estudo ou que abandonam o tratamento são exemplos de indivíduos com tempo de vida censurado.

### **2.1.1 Censura**

A censura pode ser registrada em um intervalo de tempo, ou seja, tem um início que é a entrada do indivíduo no estudo e durante esse acompanhamento pode ocorrer o evento de interesse. Existem três tipos de censura: censura à direita, censura à esquerda e censura intervalar.

#### *2.1.1.1 Censura à Direita*

Essa análise, acontece pelo motivo dos indivíduos não quererem permanecer no estudo e por isso a informação fica incompleta; ou porque o indivíduo não experimentou o evento ou a morte antes do fim do estudo; ou até mesmo que o contato foi perdido com este (BORGES, 2014, p.06).

É aquela, em que o evento de interesse não acontece até o fim do estudo. São às vezes censurados por falha ou censurados por tempo. O qual o tempo inicial é melhor que o tempo de análise. Onde “ [...] o tempo de ocorrência do evento está à direita do tempo de interesse, ou seja, o tempo entre o início do estudo e o evento é maior do que o tempo observado. ” (CUNHA, 2014, p.02). Um indivíduo está em observação, neste caso ele utiliza as informações nesse tempo sem a ocorrência do evento de interesse.

Portanto, a censura à direita são observações exatas. Por exemplo, suponha que 4 pessoas apresentaram uma doença na medula óssea, onde foram analisados os tempos para o desenvolvimento de uma infecção que é o evento de interesse, nesse caso o pesquisador tem que mostrar os tratamentos realizados e concluir a análise antes do fim do estudo. Existem três tipos de censura à direita, sendo elas: *censura do tipo I*, *censura do tipo II* e *censura aleatória*.

#### 2.1.1.1.1 Censura do tipo I

O estudo só será concluído após um tempo determinado para o fim do estudo, onde esse tempo pode variar de indivíduo a indivíduo. Entram no estudo em períodos de tempos diferentes onde o tempo de censura para cada indivíduo é observado assim que entra no estudo, no qual cada indivíduo tem estabelecido o seu tempo de censura. As análises do evento de interesse não são analisadas até esse tempo, ou seja, são censuradas.

#### 2.1.1.1.2 Censura do tipo II

Conhecida como censura por falha quer dizer que os indivíduos entram ao mesmo tempo e a análise só é realizada até que ocorra as primeiras falhas. O qual pode ser preciso um tempo longo para que todos os eventos falhem. De acordo com Aguilar (2009) em sua tese na qual ele utiliza a censura tipo II como exemplo no estudo dos 100 ratos em que seu evento principal é a morte, definindo 3 tipos de falhas. O primeiro evento acontece quando 15 ratos morrem durante a análise e os retiram do estudo, 15 dos 85 que sobreviveram, seguindo os 70 ratinhos no estudo. Obtendo 15 eventos e 15 censuras. Na segunda falha ocorre quando dos 70 ratinhos, morrem 20 restando 50 ratos em que retiram 10, continuando no estudo 40 ratinhos. Tendo 20 eventos e 10 censuras. E já na terceira falha ocorre quando dos 40 ratinhos morrem 30, restando 10 ratinhos que sobreviveram. Apresentando 30 eventos e 10 censuras.

#### 2.1.1.1.3 Censura Aleatória

Acontece quando a observação for excluída durante o estudo antes de ter ocorrido o evento de interesse ou por outra causa que não se refere ao estudo, como morte acidental, migração de um indivíduo para outra cidade, desistência do paciente, etc.

#### 2.1.1.2 Censura à Esquerda

Cunha (2014) afirma que é aquela quando não conhecemos o momento que ocorre o evento de interesse, mas que o evento já aconteceu

antes do indivíduo ser analisado no estudo. Deste modo a duração do tempo estudado é maior que a do evento.

### 2.1.1.3 Censura Intervalar

Em análise ocorre quando os indivíduos são acompanhados de tempo em tempo, respeitando intervalos. Entende-se que o evento de interesse acontece em algum intervalo de tempo, isto é, em intervalos de tempos iguais.

## 2.2 Função de Sobrevivência $S(t)$

Definida como “[...] a probabilidade que um indivíduo sobreviva além de  $T$ .” (CESÁR, 2015, p.03) ou seja, uma pessoa vive mais do que o tempo determinado de  $t$  onde não experimenta a falha.

$$S(T) = P(T > t) = 1 - P(T \leq t) = 1 - F(t).$$

Em que, “ $T$ ” é uma variável aleatória que significa o tempo de sobrevivência de um indivíduo e “ $t$ ” algum valor de interesse para  $T$ .

A função de sobrevivência não é crescente, ou seja, à medida que  $t$  cresce os valores de  $S(t)$  decresce.

Para  $t = 0 \rightarrow S(t) = S(0) = 1$ ;

Para  $t = \infty \rightarrow S(t) = 0$ , portanto  $0 \leq S(t) \leq 1$ .

Segundo Aguilar (2009), a função de sobrevivência é a integral da função densidade de probabilidade considerando  $T$  uma variável aleatória contínua, ou seja,

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt.$$

Por conseguinte, a função de distribuição acumulada  $F(t)$  “é a probabilidade de uma observação não sobreviver até o tempo  $t$ ” (FERREIRA, 2014. p.13)

$$F(t) = 1 - S(t)$$

### 2.2.1 Função densidade de probabilidade ( $f(t)$ )

É a probabilidade de falha em um pequeno intervalo por unidade de tempo (CESÁR, 2015. p.04). Portanto é uma variável aleatória contínua e positiva.

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$

### 2.3 Função de risco ( $h(t)$ )

É uma função que tem importância para os dados de sobrevivência, também conhecida por taxa de falha, força de mortalidade, força de infecção, dentre outros. "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  $t_1$ , dividido pelo comprimento do intervalo". (CUNHA, 2014.p.04).

Pode ser expressa como:

$$\begin{aligned} P[t_1 \leq T \leq t_2] &= \\ &= P(T < t_2) - P(T < t_1) \\ &= 1 - P(T \geq t_2) - [1 - P(T \geq t_1)] \\ &= P(T \geq t_1) - P(T \geq t_2) \\ &= S(t_1) - S(t_2) \end{aligned}$$

Assim, a taxa de falha no intervalo de tempo  $(t_1, t_2)$  é expressa por:

$$\frac{P(t_1 \leq T \leq t_2 | T > t_1)}{(t_2 - t_1)} = \frac{P(t_1 \leq T < t_2)}{(t_2 - t_1)P(T > t_1)} = \frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)}$$

Considerando agora o intervalo  $(t, t + \Delta t)$  em que  $\Delta t > 0$  a expressão acima pode ser reescrita como:

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

Se fizermos

$$\Delta t \rightarrow 0$$



A expressão, será definida:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

Concluindo que a taxa de falha pode ser expressa como

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)} \\ &= - \lim_{\Delta t \rightarrow 0} \frac{1 - F(t + \Delta t) - (1 - F(t))}{\Delta t} \frac{1}{S(t)} \\ &= - \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} \frac{1}{S(t)} \\ &= - \frac{S'(t)}{S(t)} = - \frac{d \log S(t)}{dt} \end{aligned}$$

Como também pode ser definida

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1 - F(t) - [1 - F(t + \Delta t)]}{\Delta t S(t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t S(t)} \\ &= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{f(t)}{S(t)} \end{aligned}$$

À medida que a probabilidade de sobrevivência diminui o risco aumenta e vice-versa. A função de risco acumulado  $H(t)$  “mede o risco de ocorrência do acontecimento de interesse até ao instante  $t$ ”. (BASTOS e ROCHA, 2006. p.186). É dada por:

$$H(t) = \int_0^t h(u) du$$

Pode ser obtida a partir da função de sobrevivência

$$h(t) = - \frac{d \log S(t)}{dt}$$

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

$$f(t) \rightarrow h(t) = \frac{f(t)}{S(t)}$$

$$f(t) = h(t)S(t)$$

$$f(t) = h(t) \exp \left\{ - \int_0^t h(u) du \right\}$$

$$S(t) = \exp \{ -H(t) \}$$

$$H(t) = - \log S(t)$$

Portanto, temos quatro funções de risco, sendo elas: tipo crescente quando o risco cresce com o tempo; tipo decrescente quando o risco diminui com o tempo; tipo constante acontece em um prazo mais curto em qualquer valor do tempo e pôr fim do tipo banheira que retrata o comportamento do tempo de vida dos indivíduos em um período inicial decrescente e no período final crescente.

## 2.4 Estimador de Kaplan Meier

O estimador de Kaplan Meier é também conhecido como " limite do produto " e é um estimador não paramétrico, ou seja, não admite qualquer função de probabilidade. O método é utilizado para calcular o tempo exato de sobrevivência de cada paciente na amostra e não agrupar os tempos de sobrevivência em intervalos. Assim, esse estimador dá a ideia do tempo que ocorrem os eventos observados (FERNÁNDEZ,1995).

A variável estudada é a diferença do tempo que segue até que o evento (morte, doença, etc.) aconteça, isto é, a diferença entre a data do começo da observação e a data da ocorrência do evento de interesse.

- **Censurado**

Seja  $t_1 < t_2 < \dots < t_m$  os tempos onde ocorreram os eventos;

$Y_i(t) = 1$ , se a pessoa  $i$  está em risco no tempo  $t$  e 0 caso contrário;

$R(t_i)$  é o total de pessoas a risco no tempo  $t_i$ .

A cada tempo  $t_i$  em que houver um evento, a probabilidade de sobrevivência será o número dos que sobreviveram até aquele tempo ( $R(t_i) - N(t_i)$ ) sobre os que estavam em risco naquele tempo  $R(t_i)$ .

O estimador de Kaplan-Meier é dado por:

$$\hat{SKM}(t) = \frac{R(t_1)-N(t_1)}{R(t_1)} \times \frac{R(t_2)-N(t_2)}{R(t_2)} \times \dots \times \frac{R(t_m)-N(t_m)}{R(t_m)}$$

Ou na forma de produto:

$$\hat{SKM}(t) = \prod_{t_i \leq t} \frac{R(t_i)-N(t_i)}{R(t_i)}$$

Onde,  $N(t_i)$  é o número de elementos que falharam no tempo e  $R(t_i)$  o número de elementos que estão em risco no tempo  $t_i$ . Portanto, o método de Kaplan Meier é indicado para amostras grandes além de ser o estimador mais utilizado da função de sobrevivência, pois ele analisa as falhas no instante que ocorreram, pontualmente.

A variância do estimador Kaplan Meier é dada por:

$$Var(\hat{SKM}(t)) = (\hat{SKM}(t))^2 \sum_{j:t_j \leq t} \frac{R(t_j)-N(t_j)}{R(t_j)}$$

Assumindo erro  $\alpha$ , o intervalo de confiança é:

Limite inferior:

$$\hat{SKM}(t) - Z_{\frac{\alpha}{2}} \sqrt{var(\hat{SKM}(t))}$$

Limite superior:

$$\hat{SKM}(t) + Z_{\frac{\alpha}{2}} \sqrt{var(\hat{SKM}(t))}$$

Entretanto, o intervalo permite valores negativos e maiores que 1.

## 2.5 Teste de Log-rank

É um método escolhido para comparar as curvas de sobrevivência de duas ou mais amostras com dados sujeitos a censura. O teste de log-rank tem a hipótese nula de que não existe diferenças de sobrevivência entre os grupos. Distribuição esperada de eventos igual em todos os grupos:

$$E_k(t) = N(t) \frac{R_k(t)}{R(t)}$$

- Teste log-rank para dois grupos ( $k = 2$ ):

$$\log - rank = \frac{(O_1 - E_1)^2}{Var(O_1 - E_1)}$$

Onde,  $O_1$  é o número total de eventos observados no grupo 1 e  $E_1$  o número total de eventos esperados no grupo 1. A Variância que entra no cálculo como um fator de padronização, tem a fórmula (para  $k = 2$ ):

$$Var(O_1 - E_1) = \sum_t \frac{R_1(t)R_2(t)\Delta N(t)[R(t) - \Delta N(t)]}{R(t)^2[R(t) - 1]}$$

## 2.6 Modelo de Cox

Cox (1972) desenvolveu um modelo de regressão semi-paramétrico, também conhecido como modelo de riscos proporcionais de Cox, ou regressão de Cox. É utilizado para analisar dados do tempo de sobrevivência em que a variável resposta é o tempo até a ocorrência de um evento de interesse, ajustado por covariáveis. O modelo de regressão de Cox é escrito como:

$$h(t) = h_0(t) \times \exp(B_{1x_1} + B_{2x_2} + \dots + B_{px_p})$$

Sendo  $h(t)$  a função de risco,  $h_0(t)$  o risco basal,  $X$  o vetor de covariáveis e o vetor de parâmetros das covariáveis. Note que as covariáveis tem um efeito multiplicativo no modelo.

A razão das funções de risco para dois indivíduos com covariáveis  $z_1$  e  $z_2$ , escreve-se da seguinte forma:

$$\frac{h(t; z_1)}{h(t; z_2)} = \exp(B(z_1 - z_2))$$

Isto é, não depende do tempo  $t$ . Portanto, podemos afirmar que qualquer uma das funções descritas na expressão para dois indivíduos quaisquer, as funções de riscos são proporcionais razão pela qual este modelo é considerado um modelo de riscos proporcionais, entende-se que as covariáveis possuem um efeito multiplicativo na função de risco. Durante o seu tempo de observação dos indivíduos, a influência das covariáveis na função de risco não se altera. Normalmente, o vetor de covariáveis não é nulo para a identificação da situação padrão.

O modelo de Cox proporciona boas estimativas tanto de risco acumulado como também de probabilidades de sobrevivência, comparado com as estimativas de Kaplan-Meier.

Uma extensão do modelo de Cox é chamada de modelo de Cox estratificado, utilizado apenas quando a suposição de proporcionalidade das taxas de falha é violada e o modelo de Cox não pode ser aplicado, dessa

forma a solução é estratificar os dados de modo que seja válida a suposição de proporcionalidade em cada estrato.

A análise estratificada consiste em dividir os dados de sobrevivência em  $m$  estratos, de acordo com uma indicação de violação da suposição. Expresso como:

$$\lambda(t|x_{ij}) = \lambda_{0j}(t) \exp \{x'_{ij}\beta\}$$

Para  $j = 1, \dots, m$  e  $i = 1, \dots, n_j$ , em que  $n_j$  é o número de observações no  $j$ -ésimo estrato. As funções de taxa de falha base  $\lambda_{01}(t), \dots, \lambda_{0m}(t)$  são arbitrárias e completamente não relacionadas (COLOSIMO e GIOLO, 2006).

## 2.7 Análise dos Resíduos

Com o intuito de verificar a adequação do modelo de regressão de Cox ajustado, se faz necessário uma análise dos resíduos que é um conjunto de técnicas utilizadas para avaliar a validade dos pressupostos do modelo, sendo eles: a presença de pontos aberrantes e influentes que são observações mal ajustadas, permite também avaliar a proporcionalidade das funções de risco. Para tanto, foram utilizados quatro tipos de resíduos para o modelo de Cox permitindo analisar os pontos mencionados acima, são eles: resíduos padronizados de Schoenfeld, resíduo Martingale, resíduos Deviance e por fim o resíduo Escore.

### 2.7.1 Resíduos padronizados de Schoenfeld

Colosimo e Giolo (2006) destacam que para avaliar a suposição de taxas de falhas proporcionais no modelo de Cox pode ser avaliada através da análise dos resíduos de Schoenfeld. Considere o  $i$ -ésimo indivíduo que possui vetor de covariáveis  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  e foi observado falhar, tem vetor de resíduos de Schoenfeld  $r_i = (r_{i1}, r_{i2}, \dots, r_{ip})$  em que cada componente  $r_{iq}$ , para  $q = 1, \dots, p$ , é definido por:

$$r_{iq} = x_{iq} - \sum_{j \in R_{ti}} \frac{x_{jq} \exp(x'_j \hat{\beta})}{\exp(x'_j \hat{\beta})}$$

Os resíduos são definidos para cada falha e não são definidos para censuras. Uma forma padronizada dos resíduos de Schoenfeld é frequentemente usada e é definida por:

$$s_i^* = [I(\hat{\beta})]^{-1} R_i$$

Com  $I(\hat{\beta})$  a matriz de informação observada.

Em seguida, para avaliar a suposição e riscos proporcionais avaliada pelos resíduos de Schoenfeld, é expresso por:

$$\lambda(t|x) = \lambda_0(t)\exp\{x'\beta(t)\}$$

Tem-se a restrição de que  $\beta(t) = \beta$ , sendo uma forma alternativa de representar o modelo de Cox. A restrição  $\beta(t) = \beta$  resulta na proporcionalidade dos riscos. Se  $\beta(t)$  não for constante, o impacto de uma ou mais covariáveis na taxa de falha pode variar com o tempo. Logo, se a suposição de taxas de falha proporcionais é validada, o gráfico de  $\beta_q(t)$  versus  $t$  deve ser uma linha horizontal. Grambsch e Therneau (1994) sugerem o gráfico de  $s_{iq}^* + \widehat{\beta}_q$  versus  $t$ , para  $q = 1, \dots, p$ , ou função do tempo,  $g(t)$ . Inclinação zero mostra evidências a favor da proporcionalidade (COLOSIMO e GIOLO, 2006).

### **2.7.2 Resíduos Martingale**

Quando se procura ajustar um modelo a um conjunto de dados, a avaliação desse ajuste passa pela análise de uma estatística especial, que tem a finalidade de medir a qualidade do modelo ajustado, uma vez que os resíduos são usados para identificar discrepâncias entre um modelo ajustado e o conjunto de dados, é conveniente buscar uma definição para resíduo que leve em consideração a contribuição de cada observação sobre medida de qualidade de ajuste (FACHINI, 2006).

O resíduo Martingale é baseado no processo de contagem e é dado por:

$$\widehat{m}_i = \delta_i - \widehat{H}_i(t_i|x_i)$$

Em que,  $\delta_i$  é a variável indicadora de falha e  $\widehat{H}_i(t_i|x_i)$  é a função de risco acumulada obtida pelo modelo ajustado.

Esses resíduos são vistos como uma estimativa do número de falhas em excesso observada nos dados, mas não predito pelo modelo. Os mesmos são

usados, em geral, para examinar a melhor forma funcional (linear, quadrática, etc.) para uma dada covariável em um modelo de regressão assumido para os dados sob estudo (COLOSIMO e GIOLO, 2006).

### 2.7.3 Resíduos Deviance

Pode ser obtido por:

$$D_i = \text{sinal}(\widehat{M}_i) [-2(\widehat{m}_i + \delta_i \log(\delta_i - \widehat{m}_i))]^{\frac{1}{2}}$$

Esses resíduos são uma tentativa de tornar os resíduos Martingale mais simétricos em torno de zero, facilitam, em geral, a detecção de pontos atípicos (*outliers*). Se o modelo for apropriado, esses resíduos devem apresentar um comportamento aleatório em torno de zero. Gráficos dos resíduos Martingale, ou Deviance, contra os tempos fornecem uma forma de verificar a adequação do modelo ajustado, bem como auxiliam na detecção de observações atípicas (COLOSIMO e GIOLO, 2006).

### 2.7.4 Resíduos Escore

O resíduo escore de acordo com Bessegato (2014), tem como funcionalidade verificar a existência de pontos influentes de cada observação no ajuste do modelo, após os pontos serem identificados, pode-se então analisar a importância que eles têm na análise. É utilizada quando o indivíduo sofre mais de um evento e tem como vantagem desse resíduo ser definida para todos os tempos mesmo onde não ocorre evento, melhorando assim a análise quando há muita censura, como também verifica a influência de cada observação nos parâmetros do modelo de Cox, como:

$$\Delta\beta = \hat{\beta} - \hat{\beta}(-i)$$

Logo, é a diferença entre o vetor de parâmetros estimados pelo modelo e o mesmo estimado sem o indivíduo  $i$ .

- Quanto maior o valor de  $\Delta\beta$ , maior a influência da observação.
- Observação tem pouca influência na estimativa se a diferença for igual a zero.

### 3. MATERIAL E MÉTODOS

#### 3.1. Banco de dados

Neste estudo foi utilizado o banco de dados “*Biofeedback*” são dados incluídos no pacote *coxph* no repositório do R, constituído de um total de 33 observações que são analisados por um período de 100 dias no interesse de analisar o tempo ao início do tratamento até o paciente atingir a reabilitação completa da deglutição, em que o tempo é a variável resposta.

Foram utilizados dois tipos de tratamentos, um grupo de pacientes que receberam tratamento de *biofeedback* videoendoscópico, e outro grupo que recebeu o tratamento conservador incluindo estimulação térmica com gelo e exercícios para os lábios, língua, fechamento e elevação de laringe. Para a análise destes dados, aplicou-se os métodos não-paramétricos de Kaplan-Meier para estimar a probabilidade de sobrevivência e de log-rank escolhido para comparar as curvas de reabilitação. Também foi aplicado o modelo de Cox caracterizado pelos coeficientes  $\beta$  que medem os efeitos das covariáveis sobre a função de risco.

Foi utilizado o método *Stepwise* com o intuito de identificar qual o melhor modelo a ser considerado. Esse método tem como objetivo de remover a variável menos significativa durante cada etapa. Também foi realizado uma análise de resíduos com intuito de checar se houve violação na suposição dos riscos proporcionais.

Todas as análises foram feitas via *Software R* (R Core Team, 2021) na versão 4.1.0. Foram utilizados os pacotes *survival* Therneau (2021) para o estimador de Kaplan Meier e o teste de Log Rank, outro pacote utilizado foi o de *coxphw* Dunkler et al. (2018) para o modelo de regressão de Cox e a análise de resíduo de Schoenfeld.

### 4 RESULTADOS

No intuito de estimar a probabilidade de sobrevivência para cada tratamento observado foi aplicado o método não paramétrico de Kaplan Meier.

É possível observar na Tabela 1 que os pacientes que receberam o tratamento conservador em 9 dias apresentaram uma probabilidade de



reabilitação de 93%. Já os pacientes com mais de 50 dias de tratamento, a probabilidade de reabilitação reduz para 47%. Podemos assim observar que quanto mais tempo o indivíduo passa em tratamento, menor a probabilidade de reabilitação.

Pelos resultados mostrados na Tabela 2 para o tratamento *biofeedback*, podemos observar que em 5 dias a probabilidade de reabilitação para os pacientes são de 95% e com 33 dias de tratamento a probabilidade dos indivíduos caem para 31%.

Abaixo tem-se as estimativas de recuperação dos pacientes sob os dois tratamentos com os respectivos intervalos de confiança aos quais estão apresentadas nas tabelas 1 e 2.

**Tabela 1** - Estimativas de Kaplan-Meier para o tempo de recuperação em paciente com o tratamento Conservador (bfb=0). As siglas E.P. significa erro padrão e IC intervalo de confiança.

Tempo	N.risco	N.evento	Sobrevivência	E.P	IC 95%	IC 95% Superior Inferior
9	14	1	0,929	0,0688	0,8030	1,000
18	13	1	0,857	0,0935	0,6921	1,000
24	12	1	0,786	0,1097	0,5977	1,000
25	11	2	0,643	0,1281	0,4351	0,950
32	8	1	0,563	0,1349	0,3515	0,900
58	6	1	0,469	0,1413	0,2596	0,846
84	5	1	0,375	0,1407	0,1797	0,783
85	4	1	0,281	0,1332	0,1112	0,711
89	3	1	0,188	0,1172	0,0551	0,639

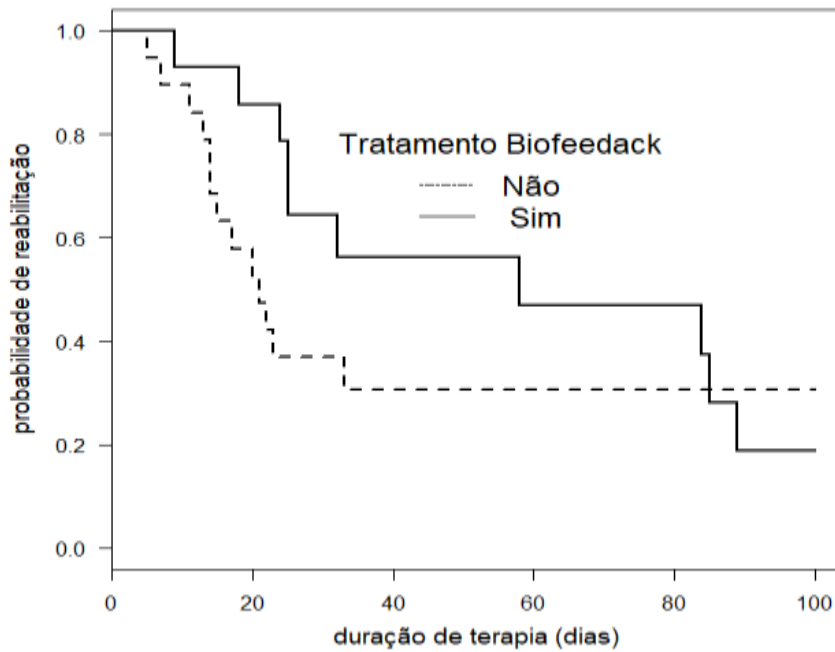
**Tabela 2** - Estimativas do Kaplan-Meier para o tempo de recuperação em paciente com o tratamento *Biofeedback* (bfb=1). As siglas E.P. significa erro padrão e IC intervalo de confiança.

Tempo	N.risco	N.evento	Sobrevivência	E.P	IC 95% Inferior	IC 95% Superior
5	19	1	0,947	0,0512	0,852	1,000
7	18	1	0,895	0,0704	0,767	1,000
11	17	1	0,842	0,0837	0,693	1,000
13	16	1	0,789	0,0935	0,626	0,996
14	15	2	0,684	0,1066	0,504	0,929
15	13	1	0,632	0,1107	0,448	0,890
17	12	1	0,579	0,1133	0,395	0,850
20	11	1	0,526	0,1145	0,344	0,806
21	10	1	0,474	0,1145	0,295	0,761
22	9	1	0,421	0,1133	0,249	0,713
23	8	1	0,368	0,1107	0,204	0,664
33	6	1	0,307	0,1079	0,154	0,611

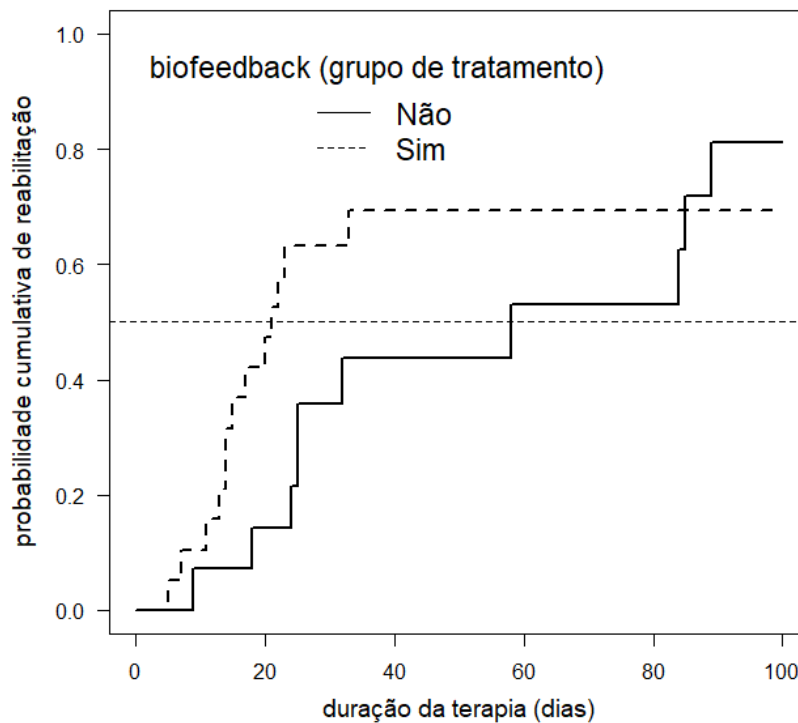
Na Figura 1, temos as curvas de reabilitação para os tratamentos obtidos pelo estimador de Kaplan Meier, em que o eixo *X* é a duração da terapia durante 100 dias e no eixo *Y* a probabilidade de reabilitação. É visto que os pacientes que tiveram o tratamento *biofeedback* tem uma probabilidade de recuperação maior comparado com os pacientes que não receberam esse tratamento.

Logo em seguida, na Figura 2, temos as probabilidades estimadas acumuladas para os tratamentos. Observa-se que foi traçado a linha reta só para verificar que 50% dos pacientes que estão no tratamento *feedback* vão obter uma probabilidade de recuperação a partir dos 30 dias de terapia, ao contrário do tratamento conservador.

**Figura 1-** Curva de Kaplan Meier para os tratamentos



**Figura 2-** Representação gráfica das estimativas de Kaplan-Meier acumulada para a função de sobrevivência para os tratamentos.



Na Tabela 3 é possível verificar que não existe uma diferença entre as curvas de sobrevivência e para isso foi realizado o teste de log-rank com o objetivo de verificar estatisticamente se as curvas de reabilitação com os respectivos tratamentos parecem apresentar alguma diferença. O teste resultou em um valor  $p$  igual a 0,5 ou seja, não rejeitando a hipótese  $H_0$ . Assim pelo teste, não há diferenças significativas entre as curvas de tratamentos (Figura 1).

**Tabela 3-** Teste de Log-Rank para os tratamentos em estudo

	N	Valor Observado	Valor Esperado	Log- Rank
Tratamento Conservador	14	10	11,8	0,549
Tratamento <i>Biofeedback</i>	19	13	11,2	0,549

Dando sequência aos resultados aplicou-se o modelo de regressão de Cox. Inicialmente foi ajustado um modelo com as covariáveis: tratamento, tempo transformado e identificação do paciente com a ajuda do comando *cluster* com intuito de agrupar a variável “identificação do paciente”. Na primeira etapa foi feito um *stepwise* para verificar qual melhor modelo e foi visto que as covariáveis tratamento e tempo transformado apresentaram melhor modelo para seguirmos com a análise.

Após isso, foi aplicado o teste de proporcionalidade no modelo ajustado mostrado na Tabela 4. De acordo com o valor  $p$  encontrado foi menor ao nível de 5% de significância, logo concluímos que há evidências de violação de proporcionalidade, ou seja, esse modelo não se deve ser aplicado.

**Tabela 4-** Teste de proporcionalidade das funções de risco no modelo ajustado.

Covariáveis	Rho( $\rho$ )	$x^2$	Valor p
Tratamento	-0,540	6,23	0,01259
Tempo Transformado	-0,263	3,51	0,06105
Global	NA	10,63	0,00491

Em seguida, foi feita a estratificação sem interação na covariável tratamento (Tabela 5) e observou-se que os pacientes no tempo, desde a cirurgia até o início da terapia, para cada acréscimo de uma unidade neste tempo, os pacientes têm 41% menos chance de reabilitação completa da deglutição, podendo variar de 5% a 63% a diminuição da chance de recuperação de acordo com seu respectivo intervalo de confiança.

Em seguida, foi realizado o teste de proporcionalidade e teve como resultado o valor  $p$  próximo ao nível de 5% de significância, sendo assim o modelo de Cox estratificado é o mais indicado para uma análise gráfica de resíduos, apresentado na Tabela 6.

**Tabela 5-** Estimativa do parâmetro referente ao modelo estratificado.

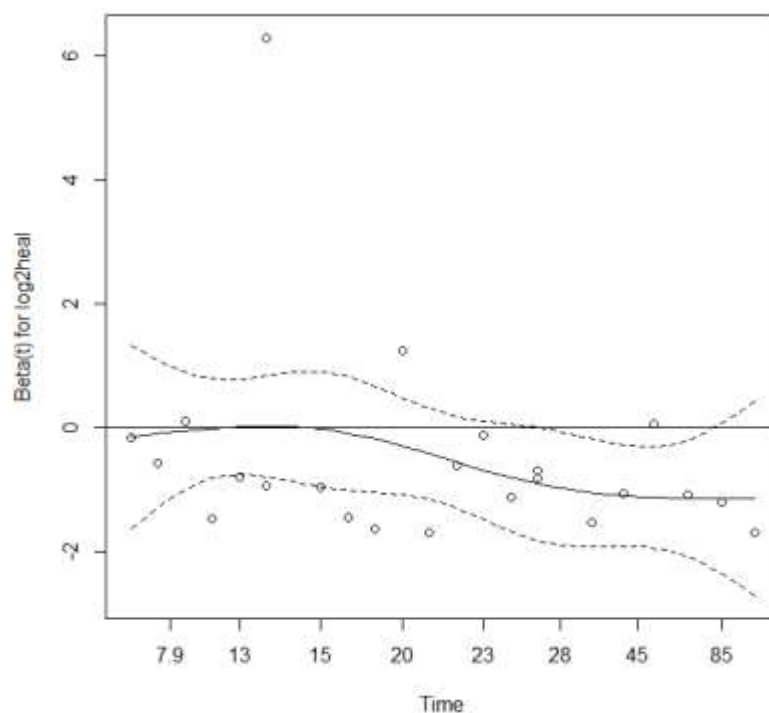
	Exp(coef)	Exp(-coef)	IC 95% Inferior	IC 95% superior
Tempo transformado	0,5993	1,669	0,3747	0,9585

**Tabela 6-** Teste de proporcionalidade da função de risco do modelo estratificado.

Covariáveis	Rho( $\rho$ )	$x^2$	Valor p
Tempo transformado	-0,269	3,26	0,0711

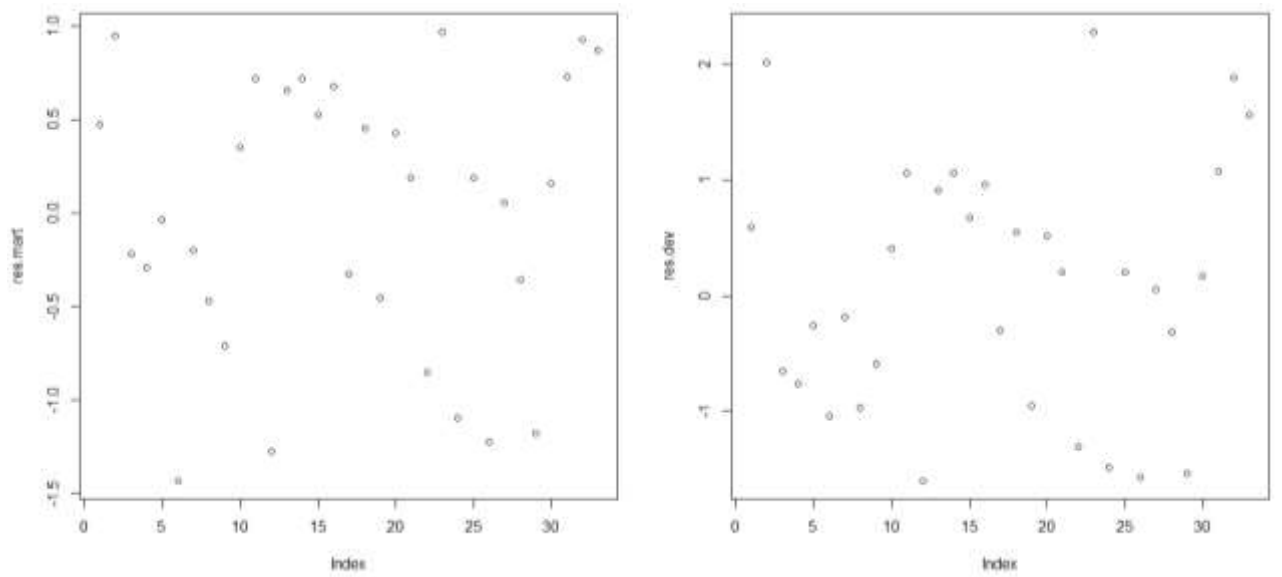
Para confirmação, foi realizada uma análise gráfica dos resíduos padronizados de Schoenfeld em que é possível notar que os resíduos possuem um padrão aleatório em torno do zero e percebe-se que há uma evidência na violação do princípio de proporcionalidade na função de risco, porém o teste da Tabela 6 não apontou essa violação. Observa-se também que, o resíduo Schoenfeld detecta indivíduos com grandes resíduos negativos, como pode ser observado na Figura 3.

**Figura 3-** Gráfico do uso dos resíduos padronizados de Schoenfeld.

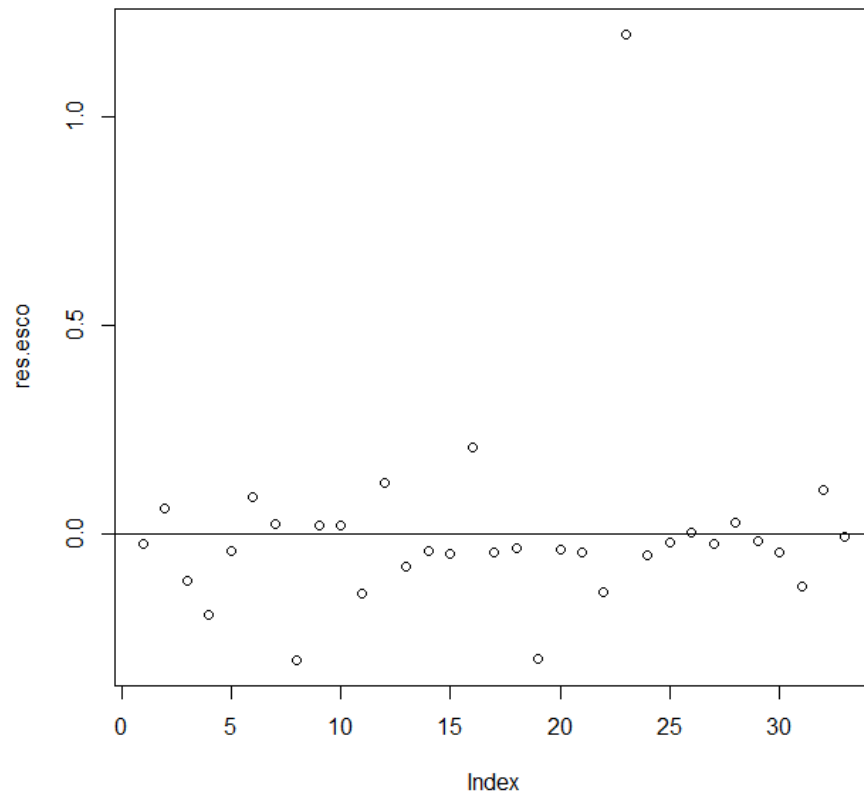


Após a verificação da proporcionalidade, procedeu-se com a análise dos resíduos Martingale que são úteis para a avaliação dos pontos aberrantes e em seguida foi realizada a análise dos resíduos Deviance, possuindo uma interpretação mais fácil se comparado com o Martingale, mostrado na Figura 4.

Nota-se que não apresentam pontos atípicos (*outliers*), pois os resíduos Martingale estão dispostos nos intervalos  $(-1.5, 1.0)$  e os resíduos Deviance dispostos nos intervalos  $(-1, 2)$ . Também é possível observar que os resíduos estão bem dispersos entre eles, não seguindo nenhum comportamento (Figura 4).

**Figura 4-** Gráfico dos resíduos Martingale e Deviance.

Por fim, foi realizado na verificação no modelo ajustado os resíduos score com o propósito de verificar a existência de pontos influentes. Analisando a Figura 5, pode-se concluir que há existência de um ponto influente no intervalo de (0,1) como também se observa que o modelo está bem adequado já que os resíduos estão bem próximos da reta.

**Figura 5-** Gráfico dos resíduos escore.



## 5 DISCUSSÃO

Com o objetivo de analisar os artigos relacionados a aplicação do modelo de Cox e as análises de resíduos, foi observado no artigo de Siqueira (2017) uma introdução ao modelo de COX com a utilização de um banco de dados que estuda o tempo até a morte de pacientes com câncer avançado de pulmão, foi mostrado que o modelo final apresentou riscos proporcionais, sendo assim, validando os pressupostos do modelo de Cox e que todas as variáveis foram significativas, como também foi realizado uma análise de resíduos de Schoenfeld confirmando a existência de proporcionalidade dos riscos para todas as covariáveis. Já no artigo de Brito *et al* (2009) foram avaliados a sobrevida de mulheres com câncer de mama no estado do Rio de Janeiro, apenas utilizaram as técnicas de Kaplan Meier e o modelo de risco de Cox tendo assim a ausência de uma análise de resíduo, já que tem como característica verificar com mais precisão o modelo ajustado.

No artigo de Cunha *et al* (2007) analisou-se o tempo até o transplante e sobrevida em pacientes com insuficiência renal crônica no Estado do Rio de Janeiro no período de 1998 a 2002. Os resultados mostraram uma aplicação da análise não paramétrica de Kaplan-Meier e a modelagem semiparamétrica com o modelo de riscos proporcionais de Cox. Foi visto que não foi aplicado uma análise de resíduos no modelo ajustado, podendo assim concluir que na maioria dos estudos realizados na área da saúde é comum visualizar artigos com resultados concluídos até o modelo de regressão de Cox. Portanto, é de suma importância realizar uma análise de resíduos para verificar as premissas de riscos proporcionais como também analisar as observações mal ajustadas do modelo, como pontos aberrantes e influentes, para assim, ter mais precisão do modelo com as variáveis estudadas.

## 6 CONCLUSÃO

Neste estudo foi realizada uma análise com pacientes após uma cirurgia de cabeça e pescoço e foi visto que houve violação na suposição dos riscos proporcionais. Foram aplicados os principais conceitos básicos da análise de sobrevivência entendendo melhor o evento estudado, principalmente como ele se relaciona com diversas covariáveis.

A curva de Kaplan Meier é uma ótima ferramenta para estimar as probabilidades de sobrevivência para cada tratamento, logo foram ajustados os modelos, indicando visualmente que não existe uma diferença entre as curvas de tratamentos e para comprovação foi feito o teste Log-Rank confirmando não haver diferença significativa entre os tratamentos estudados.

Após isso, foi ajustado o modelo de Cox com a variável tratamento e tempo transformado fez-se o teste de proporcionalidade e os resíduos de Schoenfeld resultando na existência de violação de proporcionalidade dos riscos proporcionais e, para isso, foi realizado o modelo de Cox estratificado sem interação e os resultados reforçaram na violação dos pressupostos do modelo. Concluindo assim, que o modelo de Cox estratificado se torna mais indicado do que o modelo de Cox para modelar o tempo de reabilitação após o início do tratamento. Em seguida, foi realizada uma análise de resíduos que resultou na existência de um ponto influente.

## REFERÊNCIAS

- AGUILAR, A. M. G.. **Introducción al análisis de supervivencia con R**. Tese de Doutorado. 2009. Universidad Nacional Autónoma de México - Facultad de Ciencias.
- BARROS, T. V. S. **Uma introdução à análise de sobrevivência com fração de cura**. Monografia de Graduação apresentada ao Departamento de Estatística da Universidade Federal do Rio Grande do Norte. Natal, Junho/2014.
- BASTOS, J.; ROCHA, C. **Análise de sobrevivência: conceitos básicos**. Arquivos de Medicina. Faculdade de Ciências da Universidade de Lisboa. Vol.20.nº56. 2006.
- BESSEGATO, L. F. **Análise de sobrevivência aplicada à saúde**. Departamento de estatística/ UFJF, p.01-35, 2014. Disponível em: <[http://www.bessegato.com.br/UFJF/est100\\_confiabilidade/04\\_sobrevida\\_cox.pdf](http://www.bessegato.com.br/UFJF/est100_confiabilidade/04_sobrevida_cox.pdf)>.
- BRITO, C.; PORTELA, M. C.; VASCONCELLOS, M. T. L. **Sobrevida de mulheres tratadas por câncer de mama no estado do Rio de Janeiro**. Revista Saúde Pública, v.43, n.3, p.481-9, 2009.
- BORGES, A. I. M. **Análise de sobrevivência com o R**. Dissertação Mestrado. Universidade da Madeira. Portugal. Setembro, 2014. Disponível em: <<https://digituma.uma.pt/bitstream/10400.13/732/1/MestradoAlexandraBorges.pdf>>. Acesso em 18 de julho de 2021.
- CÉSAR, K. A. **Análise estatística de sobrevivência: um estudo com pacientes com câncer de mama**. Universidade Católica de Brasília. 2015.
- COX, D. R. **Regression models and life tables (with discussion)**. Journal Royal of Statistical Society. 1972.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Blucher, 2006.
- CUNHA, F. F. **Análise de sobrevivência**. Universidade Federal do Pará. Belém, 2014.
- CUNHA, C. B.; LEÓN, A. C. P.; SCHRAMM, J. M. A.; CARVALHO, M. S.; SOUZA JÚNIOR, P. R. B.; CHAIN, R. **Tempo até o transplante e sobrevida em pacientes com insuficiência renal crônica no Estado do Rio de Janeiro**, Brasil, 1998-2002. Caderno Saúde Pública, v.23, n.4, p.805-813, Abril, 2007.

DUNKLER D., PLONER M., SCHEMPER M., HEINZE G. Weighted Cox Regression Using the R Package coxphw. *Journal of Statistical Software*, 2018.

FACHINI, J. B. **Análise de influência local nos modelos de riscos múltiplos**. Dissertação de Mestrado. Escola Superior de Agricultura Luiz de Queiroz. Piracicaba, 2006.

FERNÁNDEZ, S. P. **Análisis de supervivencia**. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo. A Coruña (España) Cad Aten Primaria 1995.p. 01-14.

FERREIRA, R. S. B. **Análise de sobrevivência com erros de classificação desconhecidos**. Dissertação de Mestrado. Universidade Federal de Minas Gerais. Belo Horizonte, 2014.

RIGÃO, H. M.; MORAES, A. B.; ZANINI, R. R. **Análise de sobrevivência: Aplicação do modelo de regressão de Cox**. Revista Ciência e Natura, v.42, Commemorative Edition: Statistic, e20, 2020.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.  
SIQUEIRA, T. S. **Introdução ao modelo de Cox**. Monografia. Universidade Federal do Ceará. Fortaleza, 2017.

THERNEAU, T. **A Package for Survival Analysis in R**. R package version 3.2-11. (2021). Disponível em: <<https://CRAN.R-project.org/package=survival>>.

THERNEAU, T.M., GRAMBSCH, P.M. (2000). **Modeling Survival Data: Extending the Cox Model**. Springer-Verlag, New York.

## **AGRADECIMENTOS**

Primeiramente agradeço a Deus por me proporcionar a trilhar nos melhores caminhos.

À minha família e principalmente minha irmã, professora e coorientadora, Dra. Pollyanna Kelly de Oliveira Silva que sempre esteve comigo me apoiando e incentivando durante todo o período de curso.

A meu orientador, Dr. Tiago Almeida de Oliveira por me proporcionar os ensinamentos para o desenvolvimento desse trabalho e por aceitar minha orientação.

A todos os professores do Departamento de Estatística que foram fundamentais na construção do aprendizado.

Enfim, a todos que contribuíram para eu chegar até aqui. Obrigada!