

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

**Um estudo de modelos de riscos competitivos em
análise de sobrevivência.**

Trabalho de Conclusão de Curso

Resumo

Análises estatísticas que envolvem registros de tempos até a ocorrência de um ou mais eventos de interesse são conhecidas como análise de sobrevivência ou de confiabilidade. Ao analisar conjuntos de dados dessa natureza, um dos objetivos está em estimar a função de sobrevivência ou confiabilidade das unidades amostrais (pessoas, sistemas, componentes, etc.). Exemplos de tempos e de eventos que geram tais tempos são: tempos até que máquinas de lavar roupas parem de funcionar, tempos até a morte/recuperação de pacientes diagnosticados com certa doença, tempos até que crianças aprendam a escrever, etc. Entre algumas situações que podem ocorrer nesse tipo de estudo destacamos duas: (a) nem todas as unidades amostrais sofrem o evento de interesse (censura) e, (b) o evento pode ocorrer por diversas causas (riscos competitivos). Nos dois casos, essas informações podem ser incorporadas à análise estatística.

A partir dos registros de tempos completos, censuras, tipos de causa e covariáveis, é possível ajustar um modelo aos dados, e além de estatísticas ou funções de interesse. Neste trabalho o objetivo está em estimar a função de sobrevivência de tempos até que determinado evento ocorra, considerando riscos competitivos, sob os enfoques não-paramétrico, via estimador de Kaplan-Meier e suas modificações e, paramétrico, via modelo de Cox e o modelo proposto por Fine e Gray. Além disso, mostrar em quais situações são melhores utilizados, bem como suas vantagens e desvantagens.

Palavras-chave: Análise de sobrevivência; Estimador de Kaplan-Meier; Modelos de Cox; Modelos de Fine e Gray; Riscos competitivos.

Abstract

Statistical analyzes that involve recording times and the occurrence of one or more events of interest are known as survival or reliability analysis. When analyzing data sets of this nature, one of the objectives is to estimate the survival or reliability function of the sampling units (people, systems, components, etc.). Examples of times and events that generate such times are: times until washing machines stop working, times until death/recovery of patients presenting with a certain disease, time until children learn to write, etc. Among some situations that can occur in this type of study, we highlight two: (a) not all sampling units suffer the event of interest (censorship) and, (b) the event can occur for several reasons (competitive risks). In both cases, this information can be incorporated into statistical analysis.

From records of complete times, censoring, types of cause and covariates, it is possible to adjust a model to the data, and also statistics or functions of interest. In this work the objective is to estimate the survival function of times in which a certain event occurs, considering competitive risks, under non-parametric approaches, by the Kaplan-Meier estimate and its modifications and, parametric, by the Cox model and the model proposed by Fine and Gray. Furthermore, show in which situations they are best used, as well as their advantages and disadvantages.

Keywords: Survival analysis; Kaplan-Meier estimator; Cox models; Fine and Gray models; Competitive risks.

Lista de Figuras

2.1	Tempos completos e exemplos de tipo de censura à direita.	14
2.2	Censura à esquerda.	15
2.3	Censura intervalar.	15
4.1	Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier(EKM) e Kaplan-Meier Ponderado(EKMP).	36
4.2	Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier Ponderado Modificado(EKMPM) e de Huang.	37
4.3	Curva de sobrevivência estimada referente aos dados estudados dada pelo Estimador de Kaplan-Meier para riscos competitivos(EKMRC).	37
4.4	Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier(EKM) e Kaplan-Meier Ponderado(EKMP) considerando censura na cauda.	38
4.5	Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier Ponderado Modificado(EKMPM) e de Huang considerando censura na cauda.	39
4.6	Curva de sobrevivência estimada referente aos dados estudados dada pelo Estimador de Kaplan-Meier para riscos competitivos(EKMRC) considerando censura na cauda.	39

Lista de Tabelas

4.1.1 Estimativas de Kaplan-Meier e suas modificações.	36
4.2.1 Estimativas de Kaplan-Meier e suas modificações considerando censura maior na cauda.	38
4.3.2 Resultados do Modelo Cox.	40
4.3.1 Dados simulados para a aplicação.	41
4.3.3 Resultados do modelo de Fine e Gray para o evento de interesse(câncer de pulmão).	41
4.3.4 Resultados do modelo de Fine e Gray para o evento competitivo(outras causas).	42

Sumário

1	Introdução	9
1.1	Objetivos	11
2	Estimadores não paramétricos	13
2.1	Alguns conceitos na área de Sobrevivência	13
2.2	Estimador de Kaplan-Meier	16
2.3	Estimador de Kaplan-Meier Ponderado	18
2.4	Estimador de Kaplan-Meier Ponderado Modificado	18
2.5	Estimador de Sobrevivência de Huang	19
2.6	Estimador de Kaplan-Meier para riscos competitivos	19
3	Modelos com Riscos Competitivos	23
3.1	Introdução	24
3.2	Modelo de Cox	25
3.2.1	Ajustando o Modelo de Cox	26
3.2.2	Método de Máxima Verossimilhança Parcial	26
3.2.3	Interpretação dos coeficientes	28
3.2.4	Adequação do Modelo de Cox	28
3.3	Modelo proposto por Fine e Gray	30
3.4	Outros modelos que abordam riscos competitivos	32
3.4.1	Modelo de Vallejos e Steel	32
3.4.2	Modelo proposto por Tortorelli	32
4	Aplicação	35
4.1	Aplicação 1 - Abordagem Não Paramétrica	35
4.2	Aplicação 2 - Abordagem Não Paramétrica	38
4.3	Aplicação 3 - Abordagem Paramétrica	40

5 Conclusão	45
Referências Bibliográficas	47

Capítulo 1

Introdução

Diferentes modelos e abordagens em análise de sobrevivência ou confiabilidade têm sido utilizados em pesquisas nas áreas médicas, das engenharias, de ciências sociais, etc., com o objetivo de entender quais covariáveis afetam o comportamento da variável resposta (tempo até a ocorrência de um evento de interesse). Alguns exemplos de variáveis respostas são os registros dos tempos: (a) de cura ou morte de pacientes com determinada doença; (b) de falha de componentes eletrônicos; (c) até que clientes de uma instituição financeira se tornem inadimplentes, entre outros. Após definir como se dará o registro dos tempos observados (unidade de medida, período de estudo, etc.), o interesse está em estimar e interpretar a função de sobrevivência e/ou funções de risco para o conjunto de observações, comparar funções de sobrevivência e/ou de risco e, avaliar a relação das variáveis explicativas com o tempo de sobrevivência.

Esta área tem metodologia própria para ajuste de modelos, pois trata de uma variável aleatória positiva (tempo até a ocorrência do evento), geralmente, com comportamento assimétrico, sendo possível incorporar na análise informações parciais (censuras) e de causas que competem entre si para a ocorrência do evento (riscos competitivos). Diversos modelos podem ser usados para descrever o tempo de ocorrência (do evento) e o ajuste destes (estimação dos parâmetros) pode ser realizado sob uma das abordagens, não-paramétrica (frequentista ou Bayesiano), paramétrica (frequentista ou Bayesiano) ou ainda semi-paramétrica.

Entre as metodologias mais conhecidas nesta área estão o estimador não-paramétrico produto-limite, conhecido como estimador de Kaplan-Meier, proposto por [Kaplan e Meier \(1958\)](#) e o modelo de regressão de riscos proporcionais de Cox ([Cox, 1972](#)). Em ambas abordagens, o tempo, observado ou censurado, é determinado por um único evento. O

estimador de Kaplan-Meier é uma das técnicas mais simples para estimar a curva de sobrevivência, considerando a presença de censuras e, possivelmente, de presença de grupos, tratadas como fatores, porém covariáveis contínuas não podem ser incluídas. [Rich et al. \(2010\)](#) apresentam uma descrição bastante detalhada deste estimador.

Embora o estimador de Kaplan-Meier tenha sido proposto em 1958, ainda assim é um estimador muito importante e utilizado em pesquisas, principalmente em estudos relacionados à saúde. Segundo um estudo realizado por Usman em 2014 ([Usman et al., 2014](#)), o mesmo foi utilizado para estimar a mediana da distribuição dos tempos de sobrevivência de pacientes com câncer de mama pós recrutamento para o estudo. Variáveis como idade, ocupação, estágio da doença e resultado do tratamento auxiliaram o estudo realizado em 312 pacientes com câncer de mama. Como conclusão geral, a idade média dos pacientes com câncer de mama foi de 43,39 anos. O tempo médio de sobrevida global foi de 10 meses e, a partir dos dados, isso indica que, para pacientes com câncer de mama, a chance de viver mais do que 10 meses é de 50%. Já para as donas de casa, esse número passa de 10 para 16 meses. Os autores apresentaram mais detalhes sobre o estudo em questão.

Na abordagem semi-paramétrica, o modelo mais conhecido é o de riscos proporcionais. Sob a suposição de riscos proporcionais, em que a razão entre duas funções de riscos, para duas diferentes unidades amostrais, não varia com o tempo. Em 1972, David Cox ([Cox, 1972](#)) propôs um modelo composto por duas funções, sendo que uma função envolve covariáveis e parâmetros associados a estas covariáveis e uma que depende apenas dos tempos, chamada de função basal. Assim, como no caso do estimador de Kaplan-Meier, o modelo de Cox ainda é muito utilizado. [Ferraz e Moreira-Filho \(2017\)](#) apresentaram uma descrição mais detalhada do modelo de Cox e, [Pereira et al. \(2022\)](#) utilizaram o estimador de Kaplan-Meier, regressão de Cox e árvore de decisão condicional para mostrar que mulheres com subtipos moleculares triplo-negativos possuem um tempo de sobrevida menor em relação às mulheres com receptores hormonais positivos. Os dados estudados foram coletados de prontuários de mulheres que tiveram câncer de mama e fizeram tratamento de 2005 a 2015. Mais detalhes sobre o estudo podem ser encontrados no artigo dos autores.

Sob outro cenário, considere que as unidades amostrais sob estudo são observadas durante um intervalo de tempo. Neste período, algumas não sofrem o evento de interesse (censura), algumas sofrem o evento de interesse e outras sofrem por outra causa (diferente da causa do estudo) e estas causas competem entre si. Estas são chamadas de riscos

competitivos. Neste contexto, [Zare e Mahmoodi \(2013\)](#) apresentaram o estimador de Kaplan-Meier modificado baseado em riscos competitivos para estimar a função de sobrevivência e [Tortorelli \(2020\)](#) ajustou um modelo de riscos competitivos para dados discretos baseado na transformação log-log complementar, sob a abordagem de inferência Bayesiana. [Austin et al. \(2020\)](#) apresentaram uma revisão do uso de covariáveis que variam no tempo, no modelo de Fine-Gray com e sem riscos competitivos. Segundo [Lacny et al. \(2015\)](#), o método de estimação de Kaplan-Meier, teoricamente, superestima a incidência acumulada de um evento na presença de riscos competitivos. Os autores apresentaram uma revisão sistemática e meta-análise a fim de determinar o grau de superestimação do método de Kaplan-Meier comparado com o método de riscos competitivos na estimação da incidência de riscos, em uma aplicação na área médica.

Conceitos referentes à análise de sobrevivência e estimadores não paramétricos para a função de sobrevivência são apresentados no Capítulo 2 e, no Capítulo 3, é apresentado o modelo de [Fine e Gray \(1999\)](#), além do modelo de [Cox \(1972\)](#). O Capítulo 4 consiste em 3 aplicações que referenciam o conteúdo estudado neste trabalho e, no Capítulo 5 é apresentado uma conclusão geral do mesmo.

1.1 Objetivos

A partir do conteúdo trabalhado na disciplina de Análise de Sobrevivência e Confiabilidade (oferecida pelo DEs - UFSCar), o objetivo do TCC é estudar alguns dos modelos nos quais podem ser considerados riscos competitivos. Mais especificamente, o objetivo é estudar o estimador de Kaplan-Meier modificado para riscos competitivos e estudar o modelo no caso de riscos competitivos, por exemplo, modelo de Cox. Também, aplicar as metodologias estudadas em conjuntos de dados simulados e, se possível, em um conjunto de dados reais. A análise é realizada com apoio computacional do *software* utilizando o pacote *timereg*, entre outros pacotes referentes à Análise de Sobrevivência e Confiabilidade.

Capítulo 2

Estimadores não paramétricos

Neste capítulo são estudados e apresentados alguns estimadores não-paramétricos aplicados na área de Sobrevivência. O primeiro e principal estimador a ser estudado é o de Kaplan-Meier (Kaplan e Meier, 1958) e, em seguida, os estimadores modificados, que são baseados no de Kaplan-Meier. Também, são estudadas situações em que cada uma dessas modificações são mais eficazes, ou seja, suas vantagens e desvantagens, como por exemplo, situações que consideram censura na cauda e situações que envolvem diferentes tamanhos amostrais. Algumas discussões são apresentadas no sentido de avaliar os pontos que diferem entre os estimadores, destacando as maiores dificuldades e principais vantagens. Por fim, são mostradas algumas situações em que há a necessidade de usar o estimador de Kaplan-Meier para riscos competitivos.

2.1 Alguns conceitos na área de Sobrevivência

No cenário da área médica, considere um grupo com n pacientes sob tratamento de determinada doença, sendo acompanhados por um período. Para cada paciente i ($i = 1, \dots, n$), é registrado o tempo até a morte (causada pela doença - evento de interesse) ou a informação de censura e algumas covariáveis (por exemplo, idade, tipo de tratamento e o estágio em que este se encontra). Porém, o paciente pode ou não sobreviver durante o período de estudo. No caso em que ele sobreviver, a informação é parcial. Também, o paciente que passa por consultas médicas periódicas pode vir a falecer entre duas consultas, e, portanto, o médico não tem conhecimento exato do tempo no qual o evento ocorreu. A falta de informação sobre o tempo exato em que ocorreu o evento, ou mesmo se este não ocorreu no período de estudo, é chamada de censura.

Kalbfleisch e Prentice (1980), Colosimo e Giolo (2006) e Carvalho *et al.* (2011), entre outros, descrevem os tipos de censura e mecanismos que geram a censura. Os tipos de censura, definidos como à direita, à esquerda e intervalar, e os mecanismos do tipo I e tipo II são abordados a seguir:

- Censura à direita: quando o evento não ocorreu no período de estudo, mas sabe-se que não ocorreu antes. Exemplos comuns deste caso envolvem estudos na área da saúde;
- Censura à esquerda: quando o evento ocorreu antes do início do período de estudo. Exemplos comuns desta situação são estudos para determinar a idade que pacientes começaram a ler. Ao iniciar o estudo, existe a possibilidade de alguns pacientes já saberem ler, mas não se lembrarem quando;
- Censura intervalar: quando o evento ocorreu no período de estudo, mas por exemplo, entre duas consultas (visitas médicas, etc.) e o tempo exato de ocorrência não é conhecido. Geralmente, este tipo de censura ocorre quando pacientes de um estudo são acompanhados periodicamente.

Para melhor compreensão, as censuras à direita, à esquerda e intervalar podem ser vistas nas figuras 2.1, 2.2 e 2.3, respectivamente.

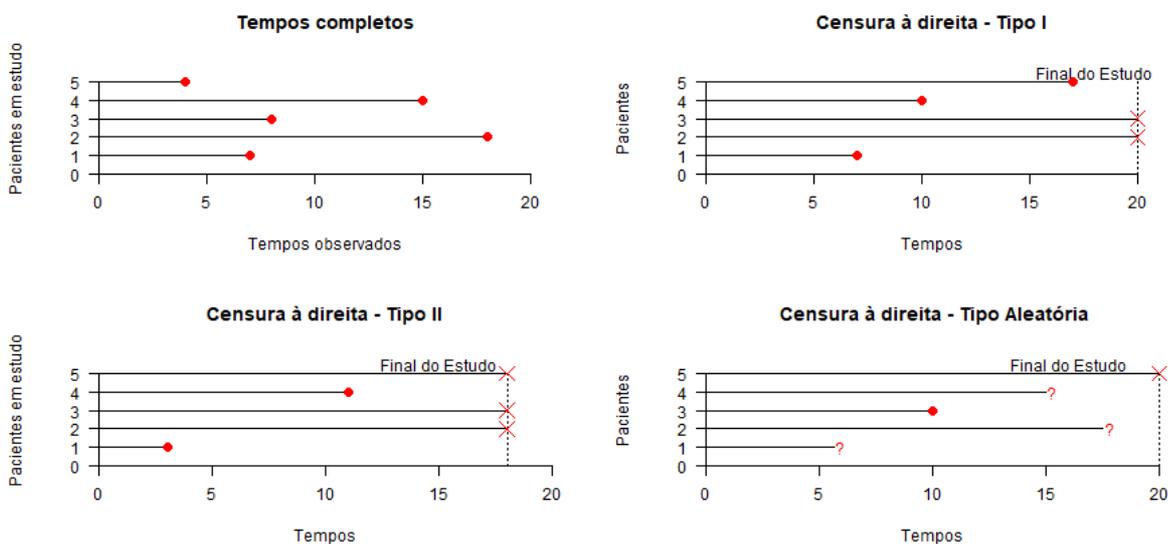


Figura 2.1: Tempos completos e exemplos de tipo de censura à direita.

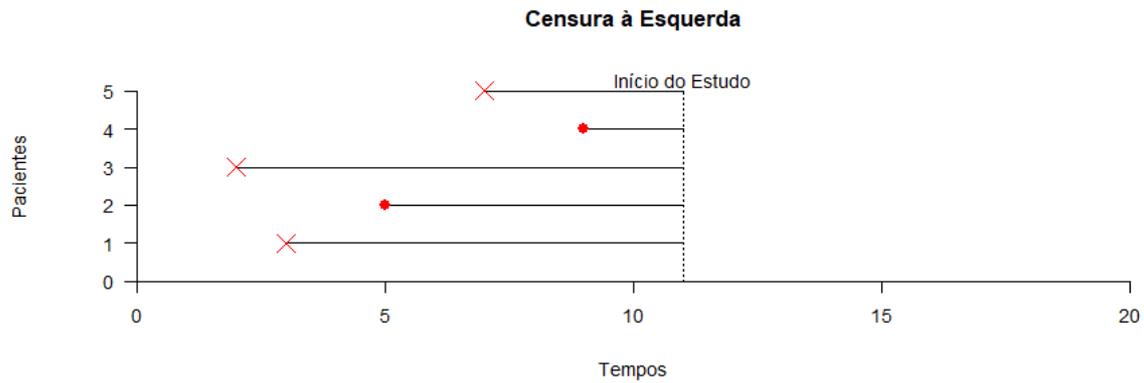


Figura 2.2: Censura à esquerda.

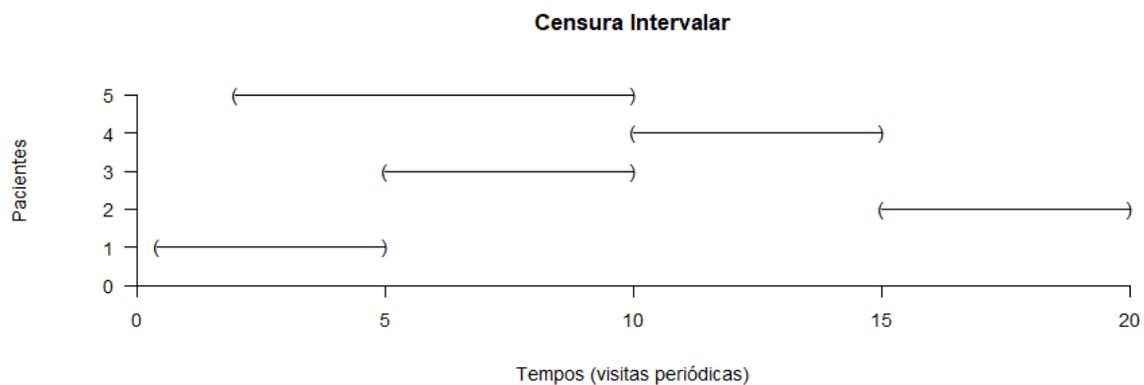


Figura 2.3: Censura intervalar.

Na Figura 2.1 podemos observar um exemplo de tempos completos, no qual 5 supostos pacientes vieram a sofrer o evento, e os 3 tipos de censura à direita. Na Figura 2.2, pode-se notar que o estudo começa no tempo 12, porém, 2 pacientes já sofreram o evento antes do início do estudo e 3 observações são censuradas. Por fim, na Figura 2.3, percebe-se que o tempo exato da censura não é informado, mas sim em qual intervalo de tempo a mesma ocorreu.

Em particular, o tipo de censura à direita é o mais comum e ocorre quando o indivíduo é acompanhado desde um tempo t_0 até um tempo pré-fixado t_f . Nesse caso, a informação extraída é somente que o evento de interesse não ocorreu entre t_0 e t_f (chamado de mecanismo do Tipo I). O mecanismo do Tipo II é comum na área industrial, em que o número de observações que sofrerão o evento é fixado (antes do início do estudo).

A censura pode ocorrer por diversas formas (tipos) e por diversas razões, por exemplo, o paciente pode sofrer o evento por outras causas, como abandonar o estudo, etc. Também, a censura pode ser informativa (perda do paciente devido à uma causa associada ao evento

de interesse) e não-informativa (quando não há evidências ou suspeitas de que o motivo da perda de informação esteja relacionado ao evento). Na metodologia da área de análise de sobrevivência (ou confiabilidade), a falta de informação sobre o tempo exato de ocorrência do evento pode ser incorporada nas análises, melhorando a qualidade dos resultados.

Considerando a abordagem não-paramétrica para estimar a função de sobrevivência, nas seções seguintes, apresentamos o estimador de Kaplan-Meier e suas formas modificadas.

2.2 Estimador de Kaplan-Meier

Como mencionado na Seção 2.1, considere n pacientes com determinada doença e sob tratamento, sendo observados por um período de tempo. Aqui, definimos o evento de interesse como sendo o falecimento (morte) do paciente. Seja T a variável aleatória que representa o tempo de vida do paciente, isto é, tempo em que ocorreu a morte contando a partir do início do estudo (ou que ele entrou no estudo). Porém, se o paciente não faleceu no período de estudo, temos uma informação parcial e, neste caso, é a censura à direita sob o mecanismo do Tipo I.

Neste contexto, a função de sobrevivência, denotada por $S(t)$, é definida como a probabilidade de sobrevivência de um paciente não sofrer o evento de interesse até o tempo t , isto é,

$$S(t) = P(T > t).$$

Esta função é não crescente tal que $S(0) = 1$ e, para $t \rightarrow \infty$, $S(t) = 0$ (ver por exemplo, (Colosimo e Giolo, 2006)).

Kaplan e Meier, em 1958, propuseram um estimador para a função de sobrevivência, baseado na contagem de pacientes que sofreram o evento de interesse e dos pacientes sob risco. Sob a abordagem não-paramétrica, a curva de sobrevivência estimada, $\hat{S}(t)$, pelo método de Kaplan-Meier (Kaplan e Meier, 1958) é dada por,

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right), \quad j = 1, 2, \dots, k, \quad (2.1)$$

em que $t_1 < t_2 < \dots < t_k$ são os tempos distintos e ordenados (em que o evento ocorreu), d_j o número de mortes no instante t_j e, n_j é o número de indivíduos que não foram censurados e ainda não sofreram o evento de interesse até o tempo imediatamente anterior

a t_j ; porém, se houver dados censurados, n_j é número de sobreviventes menos o número de casos censurados.

Na função em (2.1), destaca-se algumas vantagens e desvantagens. Entre as principais vantagens estão: este estimador leva em conta as observações censuradas, é de máxima verossimilhança (Kaplan e Meier, 1958) e é não-viciado para amostras consideradas grandes. Além disso, esse estimador é fracamente consistente e também, quando usado no caso de riscos competitivos, superestima a probabilidade de sobrevivência, fazendo com que seja necessário utilizar um estimador específico para riscos competitivos, o qual é apresentado nas próximas seções. As características de consistência e normalidade assintótica de $\hat{S}(t)$ foram provadas, sob certas condições de regularidade (ver por exemplo Breslow e Crowley (1974) e Kaplan e Meier (1958)).

A variância do estimador de Kaplan-Meier é obtida levando em consideração o erro-padrão do estimador. Greenwood (1922) apresentou uma fórmula para o cálculo da variância de $\hat{S}(t)$, dada por,

$$\widehat{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}. \quad (2.2)$$

Porém, a aplicação direta desta fórmula na construção de intervalo de confiança fornece resultados inapropriados (maiores do que um ou menores do que zero). Dessa forma, a construção de intervalos é feita usando uma solução assintótica de máxima verossimilhança por transformação logarítmica, conforme recomendado por Kalbfleisch e Prentice (1980). Assim, um intervalo aproximado com $100(1 - \alpha)\%$ de confiança ($0 < \alpha < 1$) é dado por

$$\left[\hat{S}(t) \right]^{\exp\{\pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\}}, \quad (2.3)$$

em que,

$$\hat{U}(t) = \log\left[\log(\hat{S}(t))\right], \quad (2.4)$$

com variância assintótica estimada por:

$$\widehat{Var}(\hat{U}(t)) = \frac{\sum_{j:t_j < t} \left(\frac{d_j}{n_j(n_j - d_j)}\right)}{\left[\sum_{j:t_j < t} \log\left(\frac{n_j - d_j}{n_j}\right)\right]^2} = \frac{\sum_{j:t_j < t} \left(\frac{d_j}{n_j(n_j - d_j)}\right)}{\left[\log \hat{S}(t)\right]^2}. \quad (2.5)$$

As seções a seguir são baseadas no artigo Zare e Mahmoodi (2013).

2.3 Estimador de Kaplan-Meier Ponderado

Segundo [Jan et al. \(2005\)](#), quando uma parte dos dados é censurada, o estimador de Kaplan-Meier torna-se pouco confiável e não eficiente. Uma maneira de lidar com isso, é definir um peso w_j no tempo t_j como

$$w_j = \left\{ \frac{n_j c_j}{n_j} \right\}, \quad (2.6)$$

em que $w_j = 1$ se t_j é não censurado, e $0 < w_j < 1$ se t_j é censurado e, c_j é o número de pacientes censurados no tempo t_j . A partir desta definição, o estimador de Kaplan-Meier ponderado é dado por:

$$\hat{S}_{KMP}(t) = \prod_{j:t_j \leq t} w_j \left\{ \frac{n_j d_j}{n_j} \right\}. \quad (2.7)$$

Uma das características que o estimador de Kaplan-Meier Ponderado assume, faz referência à última observação censurada. Nesse método, toda vez que a última observação for censurada, assume-se peso 0 para esta observação e, conseqüentemente, a sobrevivência estimada é igual a 0.

2.4 Estimador de Kaplan-Meier Ponderado Modificado

Uma alternativa para a característica citada na Equação (2.7) é a proposta de [Shafiq et al. \(2007\)](#) de um novo modelo, chamado de Estimador de Kaplan-Meier Ponderado Modificado. Os autores definem a função de sobrevivência da seguinte forma,

$$\hat{S}_{KMPM}(t) = \prod_{j:t_j \leq t} w_j \left\{ \frac{n_j d_j}{n_j} \right\}, \quad (2.8)$$

em que

$$w_j = \left[1 - \text{seno} \left(\frac{c_j \frac{n_j d_j}{n_j}}{n_j} \right) \right], \quad (2.9)$$

para $j = 1, \dots, n$.

É possível notar, nos estimadores dados em (2.7) e (2.8), que as formulações são muito parecidas, mudando apenas o peso atribuído para as observações, em destaque, para a última observação censurada.

2.5 Estimador de Sobrevivência de Huang

Este estimador é mais uma modificação do Estimador de Kaplan-Meier. [Huang \(2008\)](#) (ver também [\(Zare e Mahmoodi, 2013\)](#)) provou que, escolhendo pesos apropriados, seu estimador é mais eficiente que o estimador de Kaplan-Meier, tanto para dados com censura quanto para dados sem censura.

Para uma amostra com n observações e com dados censurados, os pesos (ponderações) sugeridos por Huang para primeira ($p_{n,1}$) e última ($p_{n,n}$) observações são dados por,

$$w_{1,n} = p_{n,1} = p_{n,n} = \frac{1}{2} \left(1 + \frac{n-2}{\sqrt{n(n-1)}} \right), \quad (2.10)$$

e, para as demais observações, o peso ($p_{n,i}$) é dado por,

$$w \equiv p_{n,i} = \frac{1}{\sqrt{n(n-1)}}, \quad i = 2, \dots, n-1. \quad (2.11)$$

Considerando os pesos em (2.10) e (2.11), o estimador da função de sobrevivência proposto por Huang é dado por

$$\hat{S}_H(t) = \begin{cases} 1, & t < \tau_i, \\ \prod_{i=1}^j \left(1 - \frac{d_i}{n_i} \right) \left(1 - p_{k,i} \right), & \tau_j < t < \tau_{j+1}, \quad j = 1, \dots, k-1, \end{cases} \quad (2.12)$$

sendo que τ_j , nesse caso, representa os tempos distintos e ordenados.

[Huang \(2008\)](#) afirma que se o interesse está na cauda da distribuição, o peso $w > 1/n$ deve ser escolhido a fim de ganhar eficiência na cauda de $\hat{S}_H(t)$, mas se o interesse está no meio da distribuição, então o peso $w < 1/n$ deve ser escolhido a fim de ganhar eficiência no meio do estimador $\hat{S}_H(t)$. Observe que o peso em (2.10) é maior do que $1/n$ e o peso em (2.11) é menor do que $1/n$.

2.6 Estimador de Kaplan-Meier para riscos competitivos

No contexto de dados médicos, quando nos referimos a riscos competitivos, queremos dizer que o indivíduo sob estudo pode sofrer o evento de interesse de diferentes maneiras. Para esse tipo de situação, o estimador de Kaplan-Meier não é apropriado ou indicado.

No método de Kaplan-Meier, quando um indivíduo sofre um evento de interesse por uma razão específica, supõe-se que ele possa sofrer o mesmo evento novamente por outra razão distinta. A princípio, essa probabilidade é considerada igual para todos os indivíduos. Entretanto, na prática, essa premissa não se sustenta, o que torna o estimador clássico de Kaplan-Meier para riscos competitivos ineficaz para calcular a probabilidade de sobrevivência correspondente a cada indivíduo no estudo.

No cenário em que diversas causas podem contribuir para a ocorrência do evento de interesse, [Zare e Mahmoodi \(2013\)](#) citam alguns autores que definem dois tipos de probabilidade para a obtenção da função de sobrevivência. Essas probabilidades, denominadas bruta e líquida, são definidas por,

a) bruta: probabilidade de um paciente sob estudo morrer por uma causa específica, na presença de todos os riscos que podem ocorrer com a população. Esta é dada por,

$$q_j^{\delta)} = P\{\text{indivíduo morrer no tempo } t_j \text{ pela causa } R_\delta \text{ na presença de todos os riscos presentes na população}\}.$$

b) líquida: probabilidade de um paciente sob estudo morrer se um risco específico for o único presente na população, ou ainda, a probabilidade de morte se um risco específico for eliminado da população, definida por,

$$q_j^{\delta)} = P\{\text{indivíduo morrer no tempo } t_j \text{ pelo risco } R_\delta \text{ se o mesmo for o único risco de morte}\},$$

em que δ é um indicador de causa ou risco.

Considere que existam r tipos de riscos, denotados por R_1, \dots, R_r , agindo simultaneamente em cada paciente do estudo. A relação entre as probabilidades bruta e líquida de morte, para o tipo de risco R_1 , por exemplo, é dada por

$$q_j^{1)} = \frac{q_j^{1)}}{[1 - \frac{1}{2}(q_j^{2)} + \dots + q_j^{r)}]}. \quad (2.13)$$

Além dos diferentes riscos de morte, todo tempo pode ser censurado, fazendo com que seja difícil de encontrar as taxas de sobrevivência. Como o objetivo principal é de encontrar a probabilidade de morte por um risco específico, a censura também é considerada um tipo de risco competitivo.

Seja R_r o r ésimo risco que será considerado do paciente ter seu tempo censurado. Então, para este caso, as probabilidades bruta e líquida, são dadas, respectivamente:

$q_j^{r)} = P\{\text{indivíduo morrer no tempo } t_j \text{ na presença do risco } R_r \text{ de ser censurado}\}$,
 $q_j^{r)} = P\{\text{indivíduo morrer no tempo } t_j \text{ se o risco } R_r \text{ de ser censurado for eliminado}\}$.

O estimador de Kaplan-Meier proposto para riscos competitivos é dado por

$$\hat{S}_{RC}(t) = \prod_{j:t_j \leq t} \{1 - q_j^{*1)}\}, \quad (2.14)$$

em que,

$$q_j^{*1)} = \begin{cases} q_j^{1)} & \text{se } d_j \neq 0 \\ q_j^{r)} & \text{se } d_j = 0 \end{cases}, \quad (2.15)$$

sendo que,

$$q_j^{1)} = \frac{q_j^{1)}}{[1 - 0,5(q_j^{r)})]}. \quad (2.16)$$

Como o estimador de Kaplan-Meier padrão, este estimador é uma função escada, cuja probabilidade de sobrevivência a cada morte e/ou tempo censurado é decrescente.

No próximo capítulo, são estudadas abordagens semi-paramétricas e paramétricas que envolvem riscos competitivos. As aplicações tanto da parte paramétrica, quanto da semi-paramétrica são apresentadas no Capítulo 4.

Capítulo 3

Modelos com Riscos Competitivos

Neste capítulo, é abordado uma técnica estatística importante para analisar dados de sobrevivência em situações em que indivíduos estão expostos a eventos concorrentes. O capítulo começa com uma breve introdução que apresenta o contexto em que esses modelos são aplicados e a relevância dessa análise para diversas áreas, como medicina, epidemiologia e demografia, etc.

Em seguida, é abordado o modelo de Cox, que é um dos modelos de riscos competitivos mais utilizados na literatura (Cox, 1972). O modelo de Cox é uma extensão do modelo de riscos proporcionais de Cox, utilizado para análise de dados de sobrevivência em que os eventos de interesse não são concorrentes. No contexto de riscos competitivos, o modelo de Cox é adaptado para levar em conta os eventos concorrentes, incorporando as taxas de falha associadas a cada evento.

Além do modelo de Cox, neste capítulo apresentamos outros modelos desenvolvidos especificamente para lidar com riscos competitivos. Esses modelos podem ser mais complexos, levando em consideração diferentes interações e dependências entre os eventos concorrentes.

Por fim, é enfatizado a importância de usar modelos adequados para analisar dados de sobrevivência em cenários competitivos, já que a abordagem clássica de Kaplan-Meier não é eficiente nesses casos (Carvalho *et al.*, 2011). Os modelos para riscos competitivos oferecem uma visão mais precisa e completa da dinâmica dos eventos concorrentes, permitindo uma melhor compreensão das probabilidades de sobrevivência para os indivíduos em estudo.

3.1 Introdução

Em Análise de Sobrevida e, mais especificamente, na área médica, o evento de interesse pode ocorrer devido a uma causa específica ou mesmo devido a diversas outras causas. Como introduzido na Seção 2.6, as diferentes causas de ocorrência (do evento) são denominadas eventos competitivos e introduzem o conceito de riscos competitivos. Mesmo que existam diferentes causas que concorrem entre si para que o evento ocorra, o tempo até tal ocorrência do evento de interesse e a primeira causa que ocasionou o evento são registrados. É importante destacar que na ocorrência de uma das possíveis causas, as demais não ocorrem e alteram a probabilidade de ocorrência do evento de interesse (Carvalho *et al.*, 2011).

No caso de existir apenas uma causa para a ocorrência do evento, a curva de sobrevivência estimada pelo método de Kaplan-Meier é usual e adequada para o estudo. Agora, considerando riscos competitivos, essa mesma função utilizada no método de Kaplan-Meier superestima a incidência acumulada de um evento. Isso ocorre porque a estimativa de Kaplan-Meier considera os demais tipos de sofrer o evento de interesse como censura (ver por exemplo, (Carvalho *et al.*, 2011)).

Existem diversas situações em que os modelos com riscos competitivos podem ser aplicados, mas que, na maioria dos casos, são voltados para a área da saúde. Um exemplo muito utilizado é o estudo sobre pacientes com câncer. Nesses estudos, é comum que o paciente venha a sofrer o evento de interesse por diferentes formas: recorrência da doença, desenvolvimento de metástase, trombose ou por outras causas. Outro exemplo utilizado nessa situação é o de estudos que envolvem transplante de órgãos. Nesse caso, pode ocorrer a rejeição do órgão transplantado, falha do órgão, ou a morte por outras causas, segundo da Saúde (2014). Dessa forma, a análise envolvendo riscos competitivos é bastante aplicada para estimar as taxas de ocorrência desses eventos e entender como estes interagem entre si.

Dito isso, nesta seção são apresentados modelos que podem ser aplicados especificamente para esses casos e também podem ser usados para estimar a probabilidade de um evento ocorrer, considerando outros eventos que possam afetar a probabilidade de ocorrência do evento de interesse.

3.2 Modelo de Cox

Em 1972, David R. Cox propôs um modelo de regressão semi-paramétrico, conhecido por modelo de riscos proporcionais de Cox ou modelo de Cox, ou ainda, regressão de Cox (Cox, 1972). Este modelo permite a análise de dados no caso em que a variável resposta é o tempo até a ocorrência de um evento de interesse, na presença de covariáveis e de censura (Colosimo e Giolo, 2006). O caso mais simples é dado por uma única covariável presente no estudo que representa um indicador de grupo.

Neste caso, sejam $\lambda_0(t)$ e $\lambda_1(t)$ as funções de risco do primeiro e segundo grupo, respectivamente. A função de risco representa a taxa instantânea na qual um evento ocorre em relação ao tempo, sendo muito utilizada em estudos de longevidade, medicina e outras áreas. Para o contexto do modelo de Cox, a função de risco é um componente que permite avaliar como as variáveis explicativas influenciam a probabilidade de ocorrência de um evento ao longo do tempo.

Assumindo proporcionalidade entre as duas funções, tem-se que:

$$\frac{\lambda_0(t)}{\lambda_1(t)} = K, \quad (3.1)$$

no qual K representa a razão das taxas de sofrer o evento, constante para todo tempo t .

Suponha que x seja uma variável indicadora de grupo. Então:

$$x = \begin{cases} 0, & \text{se grupo 0,} \\ 1, & \text{se grupo 1.} \end{cases} \quad (3.2)$$

e

$$\lambda(t) = \lambda_0(t) \exp\{\beta x\}, \quad (3.3)$$

ou seja,

$$\begin{cases} \lambda(t) = \lambda_0(t) \exp\{\beta\}, & \text{se } x = 1, \\ \lambda(t) = \lambda_0(t), & \text{se } x = 0. \end{cases} \quad (3.4)$$

Dessa forma, em (3.4) temos o modelo de Cox para o caso mais simples, com apenas uma covariável. De uma forma geral, seja p o número de variáveis independentes, de modo

que x seja um vetor com os componentes $x = (x_1, \dots, x_p)'$. O modelo de Cox é dada por:

$$\lambda(t) = \lambda_0(t)g(x \beta), \quad (3.5)$$

em que g é uma função não-negativa, tal que $g(0)=1$. A Equação (3.5) é composta por um termo paramétrico e outro não paramétrico. O termo não paramétrico $\lambda_0(t)$ é uma função não-negativa do tempo e é chamado de função basal, já que $\lambda(t) = \lambda_0(t)$ quando $x = 0$. Já o termo paramétrico é na maioria das vezes usado na forma multiplicativa dada por:

$$g(x \beta) = \exp\{x \beta\} = \exp\{\beta_1 x_1 + \dots + \beta_p x_p\}, \quad (3.6)$$

no qual β é o vetor de parâmetros associado às covariáveis. Esta função garante com que (3.5) seja sempre não negativa. Existem outras formas de escrever a função (3.6), porém a mais utilizada é a citada acima.

3.2.1 Ajustando o Modelo de Cox

Primeiramente, considere x o vetor de p covariáveis dado por $x = (x_1, x_2, \dots, x_p)$, β o vetor de coeficientes dado por $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ e n o número de pacientes presente no estudo.

No modelo de regressão de Cox os coeficientes β 's medem os efeitos das covariáveis sobre a função de taxa de sofrer o evento. Para que o modelo seja bom, essas quantidades devem ser estimadas a partir de uma amostra significativa, assim como em qualquer ajuste de modelos de regressão.

A presença do componente não-paramétrico $\lambda_0(t)$ faz com que o estimador de máxima verossimilhança seja ineficaz. Dessa forma, Cox (1972) apresenta uma proposta, que está em condicionar a construção da função de verossimilhança ao conhecimento do que se sabe do passado de falhas (ocorrências do evento) e censuras, eliminando, desta forma, a função de risco de base na função de verossimilhança. O método proposto é chamado de máxima verossimilhança parcial.

3.2.2 Método de Máxima Verossimilhança Parcial

Primeiramente, considere uma amostra de n indivíduos, e que existam $k \leq n$ falhas distintas, nos tempos $t_1 < t_2 \dots < t_k$. Para entender de forma simples o método de máxima

verossimilhança parcial (Colosimo e Giolo, 2006), considere o seguinte argumento:

$$\begin{aligned}
& P[\text{indivíduo sofrer um evento em } t_i \mid \text{sofrer o evento em } t_i \text{ e passado até } t_i] = \\
& = \frac{P[\text{indivíduo sofrer um evento em } t_i \mid \text{sobreviveu em } t_i \text{ e passado até } t_i]}{P[\text{sofreu o evento em } t_i \mid \text{passado até } t_i]} \\
& = \frac{\lambda_i(t|x_i)}{\sum_{j \in R(t_i)} \lambda_j} = \frac{\lambda_0(t) \exp\{x_i \beta\}}{\sum_{j \in R(t_i)} \lambda_0(t) \exp\{x_j \beta\}} \\
& = \frac{\exp\{x'_i \beta\}}{\sum_{j \in R(t_i)} \exp\{x_j \beta\}}. \tag{3.7}
\end{aligned}$$

Pode-se observar que o componente não paramétrico problemático desaparece da Equação (3.7).

Portanto, segundo Cox (1972), a função de verossimilhança utilizada para inferir os parâmetros do modelo é formada pelo produto de todos os termos representados na Equação 3.7, associados aos tempos distintos de sofrer o evento, ou seja,

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{x'_i \beta\}}{\sum_{j \in R(t_i)} \exp\{x'_j \beta\}} = \prod_{i=1}^n \left(\frac{\exp\{x'_i \beta\}}{\sum_{j \in R(t_i)} \exp\{x'_j \beta\}} \right)^{\delta_i}, \tag{3.8}$$

em que δ_i é o indicador de falha.

Um ponto a se destacar é que a função de verossimilhança parcial (3.8) assume tempos contínuos, dessa forma, não assume a possibilidade de empates nos valores observados. Empates se referem a situações em que dois ou mais eventos ocorrem simultaneamente, ou seja, no mesmo instante de tempo. Esses eventos podem ser tanto eventos de falha, quanto eventos de censura. Isso acontece devido à escala de medida de tempo, isto é, o tempo não necessariamente é medido em horas, podendo em alguns estudos ser medido em minutos, dias, meses, anos, por exemplo. Quando empates acontecem entre falhas e censuras, assume-se que a censura ocorreu após a falha, o que faz com que as observações sejam incluídas no conjunto de risco em cada tempo de falha (Colosimo e Giolo, 2006).

Considere s_i um vetor formado pela soma das correspondentes p variáveis para os indivíduos que sofrem o evento no mesmo tempo t_i ($i = 1, \dots, k$) e d_i o número de falhas neste mesmo tempo. Dito isso, a função de verossimilhança parcial que considera observações empatadas, proposta por Breslow (1972) e Peto (1972) é dada por:

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{s'_i \beta\}}{[\sum_{j \in R(t_i)} \exp\{x'_j \beta\}]^{d_i}}. \tag{3.9}$$

As propriedades assintóticas dos estimadores de máxima verossimilhança parcial são

úteis para a construção do intervalo de confiança e para testar hipóteses sobre os coeficientes do modelo. É possível utilizar as estatísticas de Wald, da Razão de Verossimilhança e Escore para fazer inferências sobre os parâmetros do modelo de Cox (Colosimo e Giolo, 2006).

3.2.3 Interpretação dos coeficientes

Observa-se pela Equação (3.5) que as covariáveis tem o efeito de acelerar ou desacelerar a função de risco. A propriedade de taxas proporcionais é útil para a interpretação dos coeficientes estimados. A razão das taxas de falha de dois indivíduos i e j que têm os mesmos valores para as covariáveis com exceção da l -ésima, é dada por:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \exp\{\beta_l(x_{il} - x_{jl})\}, \quad (3.10)$$

que é interpretado como a razão de taxas de falhas.

Um exemplo de interpretação pode ser dada de duas formas:

- Suponha, por exemplo, que x_1 seja uma covariável dicotômica indicando pacientes hipertensos. A taxa de morte entre os hipertensos é dada por: $\exp\{\beta_1\}$ vezes a taxa daqueles com pressão normal, mantendo como fixas as demais covariáveis.
- Uma interpretação considerando variáveis contínuas pode ser dada por: Suponha que em determinados dados, o efeito da variável idade é significativa e $\hat{e} = 1,1$. Neste caso, temos que, para cada acréscimo de 1 ano na variável idade, a taxa de morte aumenta em 10%.

Exemplos mais detalhados podem ser encontrados no livro de Colosimo e Giolo (2006).

3.2.4 Adequação do Modelo de Cox

Devido à presença do componente não paramétrico apresentado na Equação (3.5), o modelo de regressão de Cox é considerado bastante flexível. Porém, não se ajusta a qualquer situação e, além disso, como qualquer modelo estatístico, requer o uso de técnicas para avaliar sua adequação (Colosimo e Giolo, 2006). Uma maneira de avaliar a qualidade geral do modelo, assim como nos modelos paramétricos, é utilizar os resíduos de Cox-Snell.

Uma outra suposição básica do modelo de Cox é a de riscos proporcionais. Essa suposição é um dos pressupostos essenciais do modelo de Cox e significa que a razão das taxas de falha entre dois grupos de indivíduos deve ser constante ao longo do tempo, ou seja, as curvas de sobrevivência devem se manter paralelas. A não validação desta suposição pode causar vícios na estimação dos coeficientes do modelo, segundo [Struthers e Kalbfleisch \(1986\)](#).

Existem várias maneiras de avaliar a proporcionalidade dos riscos no contexto do modelo de Cox, algumas das principais são:

- Método gráfico descritivo: o método consiste em dividir os dados em m estratos, usualmente de acordo com alguma covariável. Em seguida, estima-se θ_j para cada estrato, utilizando o estimador proposto por [Breslow \(1972\)](#). Analisando as curvas do logaritmo de $\hat{\theta}_j \times t$, temos como conclusão que curvas não paralelas significam desvios da suposição de riscos proporcionais. Situações extremas de violação da suposição ocorrem quando as curvas se cruzam.
- Método com coeficiente dependente do tempo: o método consiste em analisar os resíduos de [Schoenfeld \(1982\)](#). Se o valor-p associado a esse teste for maior que um determinado nível de significância, isso sugere que não há evidências significativas contra a proporcionalidade dos riscos.
- Gráficos de Sobrevivência: Também é útil construir gráficos com as curvas de sobrevivência para diferentes grupos de indivíduos, especialmente se houver variáveis categóricas. Se as curvas se cruzarem ou divergirem de maneira significativa, isso pode indicar uma violação da suposição de proporcionalidade dos riscos.

É importante realizar uma análise cuidadosa da proporcionalidade dos riscos antes de interpretar os resultados do modelo de Cox, pois violações dessa suposição podem levar a conclusões distorcidas ou inválidas. Se a proporcionalidade não for atendida, deve-se considerar estratégias como ajustes no modelo, transformações de variáveis ou até mesmo a escolha de modelos alternativos.

Em geral, o modelo de regressão de Cox é muito utilizado na análise de dados de sobrevivência, já que permite incorporar, de forma fácil, covariáveis que dependem do tempo, mas, para o caso de riscos competitivos, existem modelos que atuam melhor, já que uma das principais desvantagens do modelo de Cox para riscos competitivos é que

ele assume a independência dos riscos competitivos, o que pode nem sempre refletir com a realidade.

Na Seção 3.3, é abordado um modelo alternativo ao modelo de Cox, desenvolvido para contornar algumas das limitações do modelo de Cox quando se trata de eventos competitivos.

3.3 Modelo proposto por Fine e Gray

O estimador foi desenvolvido por Gray (1988) e, posteriormente, refinado por Fine e Gray (1999), que propuseram um modelo para o risco da subdistribuição da função de incidência cumulativa. O estimador baseia-se no modelo de riscos proporcionais de Cox, mas leva em consideração a presença de eventos competitivos, como morte por diferentes causas. A principal ideia por trás do estimador de Fine e Gray é tratar os eventos competitivos como uma forma de censura informativa, em vez de ignorá-los ou considerá-los como censura independente. Isso é feito atribuindo diferentes pesos aos eventos competitivos, de acordo com a probabilidade deles ocorrerem em relação ao evento de interesse.

O estimador de sobrevivência de Fine e Gray (1999) permite estimar as curvas de sobrevivência cumulativa levando em consideração a competição entre os eventos. Além disso, é possível realizar análises estatísticas e inferências sobre as diferenças entre os grupos de interesse, como calcular os riscos relativos.

O conceito de incidência cumulativa é baseado no risco da subdistribuição de Gray (1988), para a k -ésima causa de falha., que pode ser escrita da seguinte forma:

$$\gamma_k(t|\mathbf{x}) = \gamma_{k0}(t) \exp(\beta_k \mathbf{x}), \quad (3.11)$$

em que:

- γ_k é o risco da subdistribuição;
- $\gamma_{k0}(t)$ é o risco base da subdistribuição;
- \mathbf{x} é o vetor de covariáveis do estudo;
- β_k o vetor de coeficientes do modelo.

A função de verossimilhança parcial para o modelo de subdistribuição é dada por:

$$L(\beta_k) = \prod_{j=1}^r \frac{\exp(\beta_k x_j)}{\sum_{i \in R_j} w_{ji} \exp(\beta_k x_i)}, \quad (3.12)$$

em que R_j é o conjunto de risco definido como: $R_j = \{i; T_i \geq t \text{ ou } (T_i \leq t \text{ e o indivíduo experimentou um evento de risco competitivo})\}$.

Dessa forma, o conjunto de indivíduos sob risco inclui, além dos indivíduos que não sofreram qualquer evento, aqueles que sofreram o evento competitivo, e que até a censura administrativa não saem do conjunto de risco (Carvalho *et al.*, 2011). Por outro lado, os indivíduos não contribuem de forma igual para a função de verossimilhança (3.12), assim, é necessário atribuir um peso que mude conforme o tempo de ocorrência do evento de interesse. Os pesos, calculados para todos os indivíduos em cada tempo t_j , são dados por:

$$w_{ji} = \frac{\hat{G}(t_j)}{\hat{G}(\min(t_j, t_i))}, \quad (3.13)$$

em que \hat{G} é o estimador de Kaplan-Meier da distribuição de censura (Tomaz, 2017).

Para um exemplo, considerando dois indivíduos i e j , para uma variável l e mantendo as outras constantes, a razão dos riscos da subdistribuição é dada por:

$$\frac{\gamma_i(t)}{\gamma_j(t)} = \exp\{\beta_l(x_{il} - x_{jl})\} = \exp(\beta_l). \quad (3.14)$$

Por exemplo, para variáveis dicotômicas, a razão dos riscos é a razão de um grupo atual em relação a um grupo base. Já para variáveis contínuas, refere-se ao efeito do aumento de uma unidade na covariável em questão (Tomaz, 2017).

Em resumo, o estimador de sobrevivência proposto por Fine e Gray (1999) é uma ferramenta estatística valiosa para analisar dados de sobrevivência em presença de eventos competitivos, fornecendo uma abordagem mais precisa e informativa para estimar as curvas de sobrevivência e realizar análises estatísticas em estudos de longo prazo. A popularidade do modelo de Fine e Gray também se explica, em parte, por este estar implementado no pacote `cmprsk` do software *R* (Tomaz, 2017). O nome “`cmprsk`” é uma abreviação de “Competing Risks” em inglês, que se refere aos eventos concorrentes (riscos competitivos) que podem ocorrer em um estudo de sobrevivência. O pacote é uma ferramenta estatística amplamente utilizada em análises de sobrevivência para lidar com dados censurados e eventos competitivos.

3.4 Outros modelos que abordam riscos competitivos

3.4.1 Modelo de Vallejos e Steel

O modelo de riscos competitivos proposto por [Vallejos e Steel \(2017\)](#) é baseado no modelo de riscos proporcionais de [Cox, 1972](#)) para dados discretos. O modelo é dado por:

$$\log \left\{ \frac{\lambda_r(t|\delta, B, x_i)}{\lambda_0(t|\delta, B; x_i)} \right\} = \delta_{rt} + x_i \beta_r, \quad (3.15)$$

para $r = 1, \dots, m$ e $i = 1, \dots, n$, sendo m o número de eventos que competem entre si, β_r o vetor dos coeficientes de regressão relacionado ao evento r , δ_{rt} o logaritmo do risco basal de observar o evento r com respeito ao risco de observar nenhum evento no tempo t , $\delta = \{\delta_{11}, \dots, \delta_{m1}, \delta_{12}, \dots, \delta_{m2}, \dots\}$ e,

$$\lambda_0(t|\delta, \beta; x_i) = 1 \sum_{r=1}^m \lambda_r(t|\delta, \beta; x_i), \quad (3.16)$$

o risco de no tempo t , nenhum evento ser observado.

A função de risco da causa específica é dada por:

$$\lambda_r(t|\delta, \beta; x_i) = \frac{\exp(\delta_{rt} + x_i \beta_r)}{1 + \sum_{s=1}^m \exp(\delta_{st} + x_i \beta_s)} \quad (3.17)$$

3.4.2 Modelo proposto por Tortorelli

O modelo proposto e apresentado por [Tortorelli \(2020\)](#) é baseado na versão discreta do modelo de riscos proporcionais para tempos contínuos, conhecido como log-log complementar, no qual, pode ser aplicado para tempos intrinsecamente discretos, segundo [\(Jenkins, 2005\)](#). Dessa forma, temos:

$$S(t|x) = S_b(t|x)^{\exp(x)}, \quad (3.18)$$

em que $S_b(t|x)$ representa a função de sobrevivência basal,

A Equação (3.18) pode ser escrita como:

$$1 - \lambda(t|x) = (1 - \lambda_b(t))^{\exp(x)}, \quad (3.19)$$

portanto,

$$\lambda(t|x) = 1 - (1 - \lambda_b(t))^{\exp(x)}. \quad (3.20)$$

Agora, aplicando a transformação log-log complementar, temos:

$$\log(-\log(1 - \lambda(t|X))) = \eta_t + x'\beta, \quad (3.21)$$

em que $\eta_t = \log(-\log(1 - \lambda_b(t)))$ é o log-log complementar do risco basal.

Portanto, o modelo abordado por (Tortorelli, 2020), estendendo o modelo da função acima para acomodar riscos competitivos, é dado por:

$$\log\{-\log\{1 - \frac{\lambda_r(t|\delta, \beta, x_i)}{\lambda_r(t|\delta, \beta, x_i) + \lambda_0(t|\delta, \beta, x_i)}\}\} = \delta_{rt} + x'_i\beta_r, \quad (3.22)$$

no qual, $r = 1, \dots, m$ são os m eventos que competem entre si, δ_{rt} o log-log complementar do risco basal de observar o evento r com respeito ao complementar do risco de observar todos os eventos menos o evento no tempo t , β_r o vetor de coeficientes de regressão relacionado ao evento r e $\delta = \{\delta_{11}, \dots, \delta_{m1}, \delta_{12}, \dots, \delta_{m2}, \dots\}$.

Capítulo 4

Aplicação

Neste capítulo, primeiramente, são apresentadas duas diferentes situações envolvendo os estimadores não paramétricos. Considere como fixado os quinze primeiros tempos dados no artigo [Zare e Mahmoodi \(2013\)](#) e um número pequeno de mortes que estão espalhadas aleatoriamente ao longo do tempo. Para exemplificar as variações dos diferentes estimadores estudados neste trabalho, a censura foi alterada em cada situação para comparar os estimadores entre si e um com os outros (Seções [4.1](#) e [4.2](#), respectivamente).

Além disso, na Seção [4.3](#), a partir de um banco de dados simulado, são usados os modelos de Fine e Gray para o evento de interesse e para o evento concorrente. Também, os resultados dos modelos são comparados com o modelo de Cox, que não é o mais adequado para lidar com situações que envolvem riscos competitivos. Ainda na Seção [4.3](#), são apresentados alguns comentários sobre aplicações reais realizadas no artigo de [Tomaz \(2017\)](#). A ideia, também, foi de replicar o estudo, acompanhando os resultados e comentando-os de acordo com a teoria apresentado ao longo desse trabalho, porém, o banco de dados estudado não foi divulgado. Como o trabalho de [Tomaz \(2017\)](#) é bastante informativo e agrega para o presente trabalho, este ,ainda assim, foi mencionado nesta Seção. .

4.1 Aplicação 1 - Abordagem Não Paramétrica

Nesse primeiro caso foi escolhido um número de censura pequeno e espalhada ao longo do tempo. Então, considerando os dados mostrados na Tabela [4.1.1](#), temos para a seguinte situação as probabilidades de sobrevivência estimada para cada estimador em cada tempo dadas por:

Tabela 4.1.1: Estimativas de Kaplan-Meier e suas modificações.

Tempos	Pacientes em risco	Morte	Censura	EKM	EKMP	EKMPPM	Huang	EKMRC
	15	0	0	1	1	1	1	1
1	15	1	0	0,933	0,933	0,933	0,885	0,933
2	14	0	0	0,933	0,933	0,933	0,824	0,933
3	14	0	0	0,933	0,933	0,933	0,767	0,933
5	14	0	0	0,933	0,933	0,933	0,714	0,933
6	14	0	0	0,933	0,933	0,933	0,665	0,933
8	14	0	1	0,933	0,867	0,867	0,619	0,867
9	13	1	0	0,862	0,862	0,800	0,538	0,800
11	12	0	1	0,862	0,790	0,733	0,501	0,733
12	11	0	0	0,862	0,790	0,733	0,466	0,733
16	11	0	1	0,862	0,718	0,667	0,434	0,667
17	10	0	1	0,862	0,646	0,600	0,404	0,600
18	9	1	0	0,766	0,574	0,534	0,351	0,533
21	8	0	0	0,766	0,574	0,534	0,327	0,533
28	8	0	0	0,766	0,574	0,534	0,310	0,533

No cabeçalho da coluna da Tabela 4.1.1, entende-se por:

- EKM: Estimador de Kaplan-Meier;
- EKMP: Estimador de Kaplan-Meier Ponderado;
- EKMPM: Estimador de Kaplan-Meier Ponderado Modificado;
- Huang: Estimador de Sobrevivência de Huang;
- EKMRC: Estimador de Kaplan-Meier para riscos competitivos.

A partir disso, suas respectivas curvas de sobrevivência estimadas são dadas por:

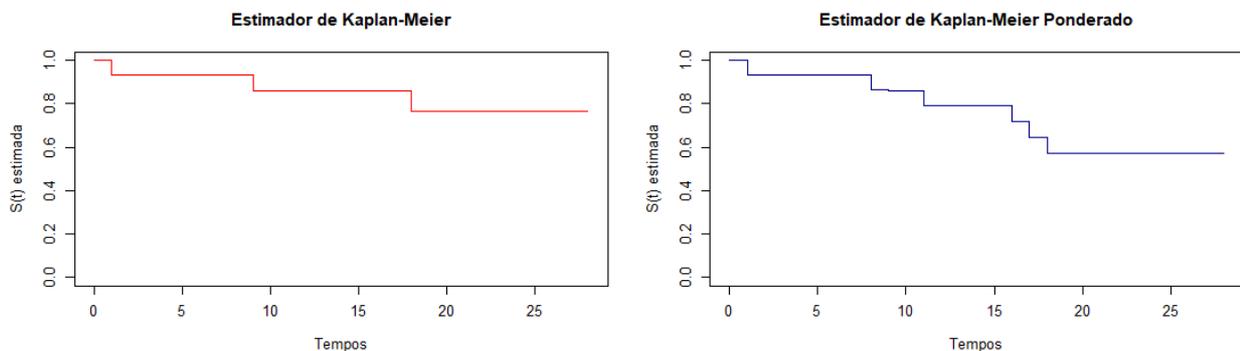


Figura 4.1: Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier(EKM) e Kaplan-Meier Ponderado(EKMP).

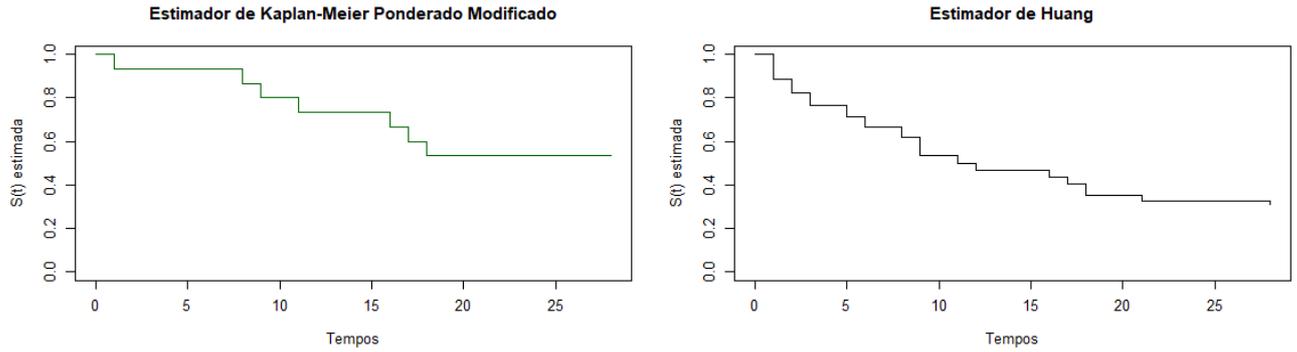


Figura 4.2: Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier Ponderado Modificado(EKMPM) e de Huang.

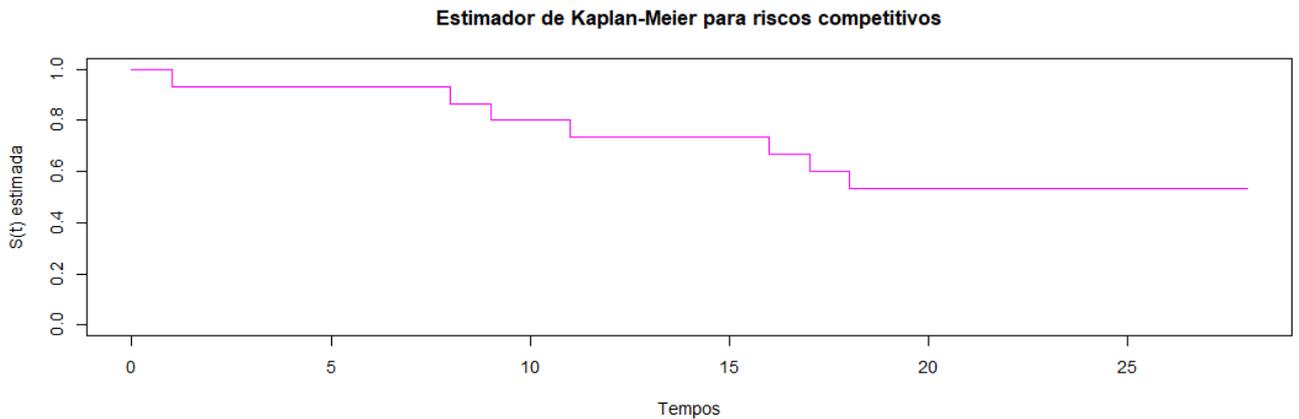


Figura 4.3: Curva de sobrevivência estimada referente aos dados estudados dada pelo Estimador de Kaplan-Meier para riscos competitivos(EKMRC).

Comparando os estimadores, percebe-se que o Estimador de Kaplan-Meier é o que, ao longo dos tempos e no final do estudo, atribui maior probabilidade de sobrevivência estimada em relação aos outros, cerca de 76%. Os estimadores de Kaplan-Meier Ponderado, Ponderado Modificado e para riscos competitivos, nessa situação de pouca censura, são bastante próximos ao longo dos tempos e, ao final do estudo, possuindo uma probabilidade de sobrevivência de aproximadamente 55%. Já em relação ao estimador proposto por Huang, podemos notar que se difere muito dos demais estudados. Como o mesmo não considera a censura no peso presente em sua função de estimação de sobrevivência, isso explica o fato da probabilidade cair mais rápido nos primeiros tempos, e terminar o estudo com probabilidade de sobrevivência próxima de 30%.

4.2 Aplicação 2 - Abordagem Não Paramétrica

Agora, para diferenciar da primeira aplicação, foi escolhido uma grande quantidade de censura presente na cauda direita. Dessa forma, podemos observar uma diferença maior entre as sobrevivências estimadas mostradas na Tabela 4.2.1.

Tabela 4.2.1: Estimativas de Kaplan-Meier e suas modificações considerando censura maior na cauda.

Tempos	Pacientes em risco	Morte	Censura	EKM	EKMP	EKMPPM	Huang	EKMRC
	15	0	0	1	1	1	1	1
1	15	1	0	0,933	0,933	0,933	0,885	0,933
2	14	0	0	0,933	0,933	0,933	0,824	0,933
3	14	0	0	0,933	0,933	0,933	0,767	0,933
5	14	0	0	0,933	0,933	0,933	0,714	0,933
6	14	0	0	0,933	0,933	0,933	0,665	0,933
8	14	0	1	0,933	0,867	0,867	0,619	0,867
9	13	1	0	0,862	0,862	0,800	0,538	0,800
11	12	0	0	0,862	0,862	0,800	0,501	0,800
12	12	0	1	0,862	0,790	0,733	0,466	0,733
16	11	0	1	0,862	0,718	0,667	0,434	0,667
17	10	0	2	0,862	0,574	0,534	0,404	0,533
18	8	1	2	0,754	0,377	0,366	0,351	0,457
21	5	0	3	0,754	0,151	0,159	0,327	0,183
28	2	0	0	0,754	0,151	0,159	0,310	0,183

O cabeçalho da coluna da Tabela 4.2.1 possui a mesma notação da Tabela 4.1.1.

As respectivas curvas de sobrevivência estimadas são dadas por:

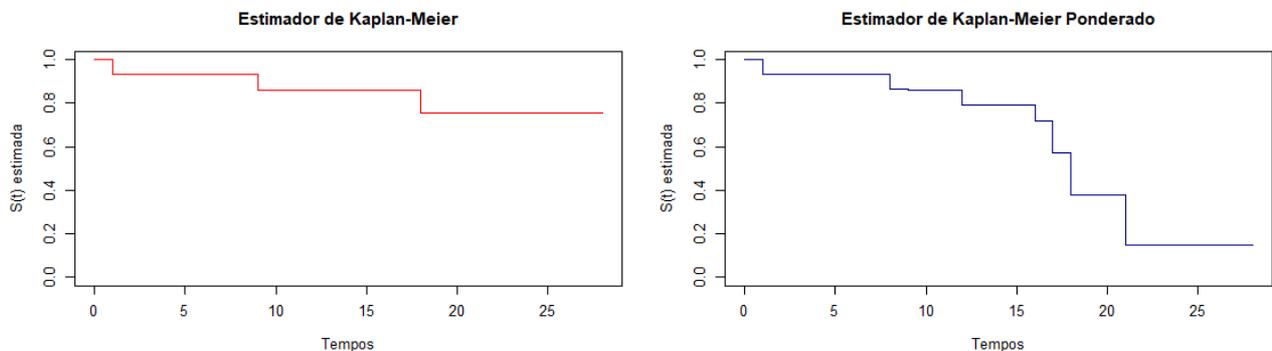


Figura 4.4: Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier(EKM) e Kaplan-Meier Ponderado(EKMP) considerando censura na cauda.

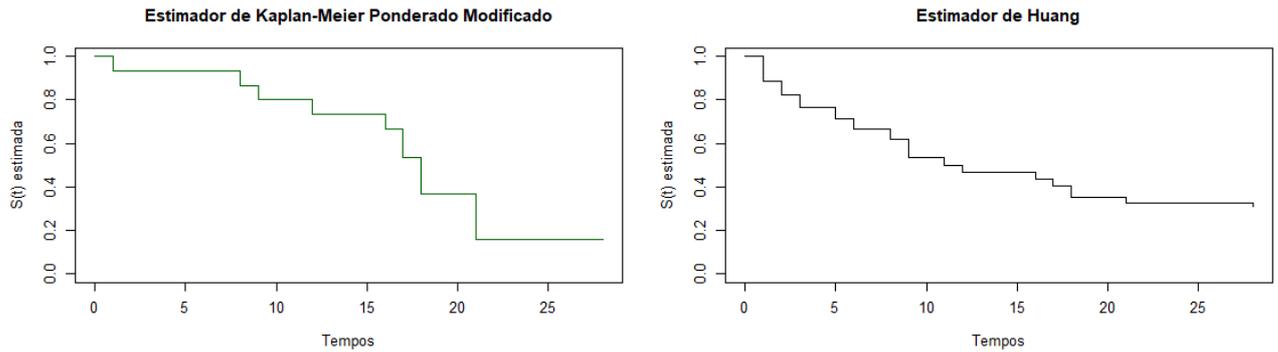


Figura 4.5: Curvas de sobrevivência estimadas referente aos dados estudados dadas pelos estimadores de Kaplan-Meier Ponderado Modificado(EKMPM) e de Huang considerando censura na cauda.

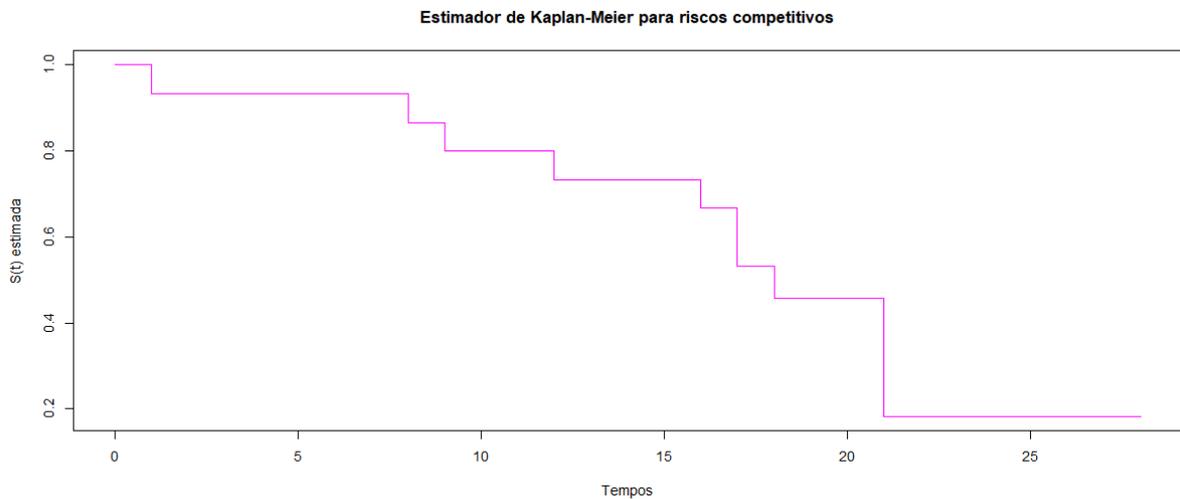


Figura 4.6: Curva de sobrevivência estimada referente aos dados estudados dada pelo Estimador de Kaplan-Meier para riscos competitivos(EKMRC) considerando censura na cauda.

Agora, percebe-se que, para casos em que a censura é maior na cauda, o Estimador de Kaplan-Meier não é eficiente. O mesmo não é ponderado pela censura, apenas a carrega para o tempo seguinte, superestimando a probabilidade na cauda. Essa afirmação fica mais perceptível quando comparada com o caso anterior, no qual as probabilidades de sobrevivência são praticamente iguais, mesmo com uma mudança brusca na quantidade de censura. Em relação aos estimadores de Kaplan-Meier Ponderado, Ponderado Modificado e para riscos competitivos, percebe-se que a probabilidade nos primeiros tempos são parecidas com a da situação apresentada na Seção 4.1, porém, pelo fato de existir uma quantidade maior de censura na cauda, as probabilidades de sobrevivência estimada caem de forma mais brusca, fazendo com que a probabilidade de sobrevivência ao final do

estudo esteja próxima dos 15%. Por fim, em relação ao estimador proposto por Huang, nota-se que não existem diferenças entre as duas situações, pelo fato de que a censura não interfere na probabilidade de sobrevivência estimada.

4.3 Aplicação 3 - Abordagem Paramétrica

O objetivo nessa aplicação é comparar duas abordagens estudadas nesse trabalho utilizando um banco de dados com a presença de eventos competitivos.

Os dados foram simulados considerando uma amostra de 20 pacientes. A simulação foi realizada no *software* . Suponha que esses 20 pacientes participam de um estudo de câncer de pulmão. Dessa forma, o evento de interesse é a morte por câncer e o evento competitivo considera a morte por outras causas. As variáveis que compõem o banco de dados são:

- tempo: tempo decorrido até o evento ou censura;
- status: 0 = censura; 1 = evento de interesse; 2 = evento competitivo;
- sexo: 0 = masculino; 1 = feminino;
- idade: idade dos pacientes (em anos);
- evento: variável indicadora do evento competitivo.

Os dados são mostrados na Tabela 4.3.1

A partir dos dados, a ideia é ajustar e comparar o modelo de Cox (Cox, 1972) e o modelo de Fine e Gray (Fine e Gray, 1999), apresentados no Capítulo 3. O modelo de Fine e Gray pode ser aplicado no *software* utilizando a função *crr* da biblioteca *cmprsk*.

Primeiramente, ajustando o modelo de Cox, obtemos o seguinte resultado:

Tabela 4.3.2: Resultados do Modelo Cox.

Variável	Coefficiente	Exp(Coefficiente)	<i>p</i> -valor
Idade	-0,03983	0,96095	0,497
Sexo	-0,61932	0,53831	0,421

Tabela 4.3.1: Dados simulados para a aplicação.

Tempo	Status	Sexo	Idade	Evento
1	2	0	74	1
3	1	0	68	0
4	2	0	56	1
5	1	1	57	0
7	2	1	60	1
7	2	0	66	1
8	1	1	70	0
9	2	1	60	1
12	1	0	59	0
12	2	1	70	1
13	1	1	74	0
15	2	0	71	1
17	1	0	76	0
25	1	0	59	0
34	0	1	67	0
40	1	1	65	0
52	1	0	71	0
63	2	1	58	1
66	1	1	61	0
71	1	1	70	0

Observando os p -valores para as variáveis Idade e Sexo, na Tabela 4.3.2, podemos concluir que ambas as variáveis não são significativas para o modelo.

Agora, para o modelo proposto por Fine e Gray para a morte por câncer de pulmão (evento de interesse), temos o resultado mostrado na Tabela 4.3.3:

Tabela 4.3.3: Resultados do modelo de Fine e Gray para o evento de interesse(câncer de pulmão).

Variável	Coefficiente	Exp(Coefficiente)	Erro Padrão	z	p -valor
Idade	-0,00678	0,993	0,6041	-0,0112	0,99
Sexo	0,02685	1,027	0,0518	0,5187	0,60

Observando a Tabela 4.3.3, pode-se notar que as covariáveis não são significativas para o modelo. Também, a partir da Tabela 4.3.3, observando o exponencial do coeficiente podemos tirar algumas conclusões a partir do modelo:

- Para a Idade: O valor encontrado é de 0,993, indicando que o aumento de uma unidade na variável idade está associado a uma diminuição de cerca de 0,7% no

risco de ocorrência do evento de interesse (câncer de pulmão), considerando todas as outras variáveis constantes.

- Para o Sexo: O valor encontrado é de 1,027, indicando que o o paciente ser do sexo feminino, está associado a um aumento de cerca de 2,7% no risco de ocorrência do evento de interesse, considerando todas as outras variáveis constantes.

Agora, para o modelo proposto por Fine e Gray para o morte por outras causas (evento competitivo), temos o resultado mostrado na Tabela 4.3.4:

Tabela 4.3.4: Resultados do modelo de Fine e Gray para o evento competitivo(outras causas).

Variável	Coefficiente	Exp(Coefficiente)	Erro Padrão	z	p -valor
Idade	-0,4551	0,634	0,6767	-0,672	0,50
Sexo	-0,0439	0,957	0,0599	-0,732	0,46

Analisando o modelo para a morte por outras causas (evento competitivo), dados pela Tabela 4.3.4, podemos dizer também que ambas as covariáveis não são significativas para o modelo (p-valores são muito altos).

Em relação ao efeitos das covariáveis, podemos dizer que:

- Para a Idade: O valor encontrado é de 0,634, indicando que o aumento de uma unidade na variável idade está associado a uma diminuição de cerca de 36,6% no risco de ocorrência do evento competitivo(outras causas), considerando todas as outras variáveis constantes.
- Para o Sexo: O valor encontrado é de 0,957, indicando que o paciente ser do sexo feminino, está associado a uma diminuição de cerca de 4,3% no risco de ocorrência do evento competitivo, considerando todas as outras variáveis constantes.

Em geral, ambos os modelos (Cox e Fine e Gray) indicam que as variáveis analisadas não têm um efeito estatisticamente significativo no risco de ocorrência do evento de interesse. Em ambos os casos, os p-valores são altos, sugerindo que as variáveis não são relevantes para a análise. Embora os modelos não sejam significativos, a ideia principal dessa aplicação, é mostrar como os modelos podem ser aplicados,, suas respectivas interpretações caso fossem significativos e comparações entre si, além de apresentar a função *crr* do *software* .

A diferença fundamental entre os modelos é que o proposto Fine e Gray leva em consideração eventos competitivos, enquanto o modelo de Cox não leva em conta esses eventos concorrentes. Portanto, se a natureza do estudo envolve a presença de eventos competitivos, o modelo de Fine e Gray pode ser mais apropriado para avaliar as variáveis em questão.

Comparando os dois modelos de Fine e Gray, pode-se tirar conclusões a partir das duas variáveis:

- Em relação à variável Idade, o modelo que considera o evento concorrente tem uma diminuição no risco de ocorrência muito maior comparado ao modelo que considera o evento de interesse, quando aumentamos a idade em uma unidade.
- Em relação à variável Sexo, para o modelo que considera o evento de interesse, o risco de ocorrência do evento é maior quando o paciente é do sexo feminino, já para o modelo que considera o evento competitivo, quando o paciente é do sexo feminino, o risco de ocorrência é menor, comparado ao sexo masculino.

Além deste exemplo simulado, podemos mencionar o trabalho de [Tomaz \(2017\)](#). A tese de doutorado “Modelagem da função de incidência cumulativa na presença de riscos competitivos em Análise de Sobrevivência” é dividida em 4 capítulos, cada um contendo uma aplicação diferente. O Capítulo 3 confronta duas abordagens diferentes para dados na presença de riscos competitivos. A primeira, faz o uso da incidência cumulativa por meio do uso do complemento do estimador de Kaplan-Meier, e na modelagem do efeito das covariáveis pelo ajuste do modelo de riscos proporcionais de Cox (forma padrão). Nessa abordagem os eventos competitivos são tratados como censura. A segunda abordagem faz o uso do modelo proposto por Fine e Gray ([Fine e Gray, 1999](#)).

Os dados utilizados referem-se a equinos da raça PSI, estudados no Centro de Tratamento em Pedro do Rio, Petrópolis, RJ. Os eventos considerados neste estudo foram lesões devido à claudicação (evento de interesse) e lesões decorrente de outras injúrias (evento competitivo).

Em geral, como conclusão, a autora cita que os resultados encontrados indicam que a estimativa da incidência para cada um dos eventos, utilizando o estimador de Kaplan-Meier, é superestimada, assim como foi apresentado na teoria ao longo deste trabalho. Outro ponto citado pela autora ([Tomaz, 2017](#)) é que a modelagem dos efeitos das covariáveis sobre o risco específico, ajustado pelo modelo de Cox, não corresponde ao efeito

sobre a incidência deste evento ajustado pelo modelo de Fine e Gray. Portanto, a abordagem baseada na análise de sobrevivência padrão quando eventos competitivos estão presentes no estudo é inadequada, a alternativa mais apropriada baseia-se na abordagem dos riscos competitivos.

No final desta Seção, além do que foi apresentado, pretendeu-se também de replicar a aplicação realizada por [Tomaz \(2017\)](#) e demonstrar os resultados adquiridos, fazendo uma comparação com a parte teórica, porém, a base de dados não foi mencionada e disponibilizada. Ainda assim, como a Tese tem bastante relação com os tópicos vistos nesse trabalho, optou-se por comentá-la e fazer uma menção à mesma.

Capítulo 5

Conclusão

A análise de sobrevivência desempenha um papel muito importante para a população. Ao estimar as probabilidades de ocorrência de eventos ao longo do tempo, como doenças, falhas em dispositivos, morte, entre outros, a análise de sobrevivência oferece informações valiosas para orientar decisões clínicas e até estratégias de negócios. Além disso, a análise de sobrevivência permite a identificação de fatores de risco, a personalização de tratamentos, a avaliação da eficácia de intervenções e a previsão da vida útil de produtos. Dessa forma, essa abordagem estatística contribui de maneira significativa para melhorar a qualidade de vida da população, possibilitando intervenções oportunas, melhores tomadas de decisão e um planejamento mais eficaz para eventos que ocorrem ao longo do tempo.

Ao longo deste trabalho, exploramos o estimador de Kaplan-Meier e suas modificações como ferramentas essenciais para analisar dados de sobrevivência, destacando suas vantagens, mas também reconhecendo suas limitações em certos contextos. Ficou nítido a importância que o estimador de Kaplan-Meier tem em Análise de Sobrevivência e Confiabilidade, dado que existem diversas modificações que foram feitas a partir dele, com o intuito de adequá-lo a situações diferentes. Em relação aos modelos que consideram riscos competitivos, foram apresentados alguns estimadores que conseguem trabalhar com eventos que competem entre si, o que, na maioria das vezes, é o que ocorre em ocasiões reais e que também são de grande importância.

No Capítulo 4, conseguimos conectar a teoria com a parte prática, mostrando como essas técnicas podem ser aplicadas e mostrando que, em alguns casos, algumas delas não são eficazes. Nas duas primeiras aplicações pudemos notar a diferença que o estimador de Kaplan-Meier e suas modificações têm entre si, com ou não a presença de censura nos dados. Já na aplicação 3, percebe-se a diferença que o modelo de Cox possui em relação

ao modelo proposto por Fine e Gray. Mesmo os modelos não sendo estatisticamente significativos, assim como mencionado na teoria, pudemos notar que o modelo de Cox não é apropriado para casos que envolvem riscos competitivos e que o ideal é utilizar o modelo proposto por Fine e Gray, que leva em conta os eventos que competem entre si.

Uma extensão deste trabalho, seria aplicar o estimador proposto por [Tortorelli \(2020\)](#), que é uma aplicação Bayesiana. Além disso, aplicar o estimador de [Fine e Gray \(1999\)](#) em um conjunto de dados reais que ainda não tenha sido trabalhado, com o objetivo de agregar de alguma forma para determinada área de pesquisa, como por exemplo, na área médica, ajudando na tomada de decisão.

O trabalho realizado foi uma experiência bastante valiosa, permitindo revisitar conceitos estatísticos básicos que havia estudado há bastante tempo na disciplina "Análise de Sobrevida e Confiabilidade". Além disso, pude aprender sobre novos estimadores mais específicos, que não haviam sido mencionados na disciplina, tanto não-paramétricos, quanto paramétricos e suas respectivas vantagens e desvantagens. Um outro ponto de bastante importância foi o aprendizado em relação à pesquisa e escrita. O trabalho exige compromisso e que as informações sejam transmitidas de maneira eficiente e organizada.

Referências Bibliográficas

- Austin, P. C., Latouche, A. e Fine, J. P. (2020). A review of the use of time-varying covariates in the Fine-Gray subdistribution hazard competing risk regression model. *Statistics in Medicine*, **39**(2), 103–113.
- Breslow, N. e Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, páginas 437–453.
- Breslow, N. E. (1972). Contribution to discussion of paper by dr cox. *Journal of the Royal Statistical Society, Series B*, **34**, 216–217.
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S. e Shimakura, S. E. (2011). *Análise de sobrevivência: teoria e aplicações em saúde*. SciELO-Editora FIOCRUZ.
- Colosimo, E. A. e Giolo, S. R. (2006). *Análise de sobrevivência aplicada*. Editora Blucher.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**(2), 187–202.
- da Saúde, M. (2014). Protocolos clínicos e diretrizes terapêuticas em oncologia.
- Ferraz, R. d. O. e Moreira-Filho, D. d. C. (2017). Análise de sobrevivência de mulheres com câncer de mama: modelos de riscos competitivos. *Ciência & Saúde Coletiva*, **22**, 3743–3754.
- Fine, J. P. e Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, **94**(446), 496–509.
- Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of statistics*, páginas 1141–1154.

- Greenwood, M. (1922). Discussion on the value of life-tables in statistical research. *Journal of the Royal Statistical Society*, **85**(4), 537–560.
- Huang, M. L. (2008). A weighted estimation method for survival function. *Applied Mathematical Sciences*, **2**(16), 753–62.
- Jan, B., Shah, S. W. A., Shah, S. e Qadir, M. F. (2005). Weighted Kaplan Meier estimation of survival function in heavy censoring. *Pakistan Journal of Statistics*, **21**(1), 55.
- Jenkins, S. P. (2005). Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*, **42**, 54–56.
- Kalbfleisch, J. e Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics. John Wiley & Sons. ISBN 9781118031230.
- Kaplan, E. L. e Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.
- Lacny, S., Wilson, T., Clement, F., Roberts, D. J., Faris, P. D., Ghali, W. A. e Marshall, D. A. (2015). Kaplan-Meier survival analysis overestimates the risk of revision arthroplasty: a meta-analysis. *Clinical Orthopaedics and Related Research* , **473**(11), 3431–3442.
- Pereira, L. C., da Silva, S. J., Fidelis, C. R., de Lima Brito, A., Júnior, S. F. A. X., dos Santos Andrade, L. S., de Oliveira, M. E. C. e de Oliveira, T. A. (2022). Cox model and decision trees: an application to breast cancer data. *Revista Panamericana de Salud Publica*, **46**:e17.
- Peto, R. (1972). Contribution to the discussion of paper by dr cox. *J. Royal stat. Soc.*, **34**, 205–207.
- Rich, J. T., Neely, J. G., Paniello, R. C., Voelker, C. C. J., Nussenbaum, B. e Wang, E. W. (2010). A practical guide to understanding Kaplan-Meier curves. *Otolaryngology—Head and Neck Surgery*, **143**(3), 331–336.
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, **69**(1), 239–241.
- Shafiq, M., Shah, S. e Alamgir, M. (2007). Modified weighted Kaplan-Meier estimator. *Pakistan Journal of Statistics and Operation Research*, **3**(1), 39–44.

- Struthers, C. A. e Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, **73**(2), 363–369.
- Tomaz, F. S. C. (2017). Modelagem da função de incidência cumulativa na presença de riscos competitivos em análise de sobrevivência.
- Tortorelli, F. A. C. (2020). *Modelos de Riscos Competitivos no Estudo de Evasão Discente*. Dissertação de mestrado, Mestrado em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística, ICMC/USP, São Carlos, S.P.
- Usman, M., Dikko, H., Bala, S., Gulumbe, S. *et al.* (2014). An application of kaplan-meier survival analysis using breast cancer data. *Sub-Saharan African Journal of Medicine*, **1**(3), 132.
- Vallejos, C. A. e Steel, M. F. J. (2017). Bayesian survival modelling of university outcomes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **180**(2), 613–631.
- Zare, A. e Mahmoodi, M. (2013). Modified Kaplan-Meier Estimator Based on Competing Risks for Heavy Censoring Data. *International Journal of Statistics in Medical Research*, **2**(4), 297–304.