



UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Deise Pereira da Silva

Introdução ao modelo de Cox com aplicação a dados de Pneus 11.00R22

Campina Grande - PB

Outubro de 2016

Deise Pereira da Silva

Introdução ao modelo de Cox com aplicação a dados de Pneus 11.00R22

Trabalho de Conclusão de Curso 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

Campina Grande - PB

Outubro de 2016

É expressamente proibida 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 da dissertação.

S586i Silva, Deise Pereira da.
Introdução ao modelo de Cox com aplicação a dados de pneus
11.00r22 [manuscrito] / Deise Pereira da Silva. - 2016.
38 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) -
Universidade Estadual da Paraíba, Centro de Ciências e
Tecnologia, 2016.

"Orientação: Prof. Dr. Tiago Almeida de Oliveira,
Departamento de Estatística".

1. Modelos semi-paramétricos. 2. Modelo de Cox. 3.
Suposição dos riscos proporcionais. 4. Análise de dados. I.
Título.

21. ed. CDD 519.53

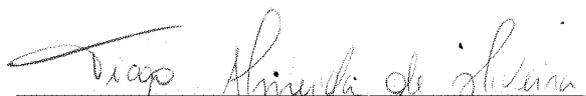
Deise Pereira da Silva

Introdução ao modelo de Cox com aplicação a dados de Pneus 11.00R22

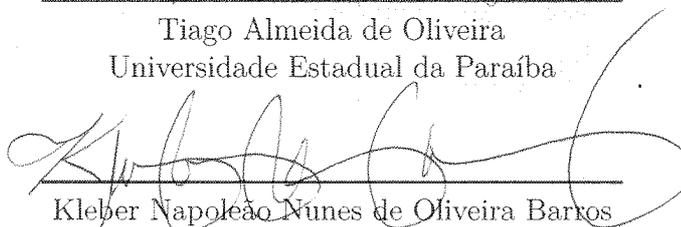
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 26 de Outubro de 2016.

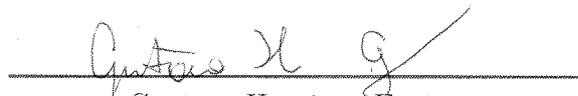
BANCA EXAMINADORA



Tiago Almeida de Oliveira
Universidade Estadual da Paraíba



Kleber Napoleão Nunes de Oliveira Barros
Universidade Estadual da Paraíba



Gustavo Henrique Esteves
Universidade Estadual da Paraíba

Em memória da minha querida avó Irací Maria Pereira, que em vida sempre acreditou em meu potencial, e com certeza ficaria muito feliz por essa realização em minha vida.

MINHA HOMENAGEM

A meus pais: Damiana José Pereira e José Severino da Silva Filho, e a meus irmãos: Jakson Pereira da Silva, Assis José da Silva, Angela Cavalcanti, que de muitas formas me incentivaram para que fosse possível a concretização deste trabalho.

OFEREÇO E DEDICO

Agradecimentos

Nenhuma batalha é vencida sozinha. Durante essa fase de evolução na minha vida, muitas pessoas importantes estiveram ao meu lado, dando motivos suficientes para que eu buscasse minha vitória e realização dos meus sonhos.

Agradeço primeiramente e acima de tudo a Deus, por me conceber o dom da vida, por me ouvir a todo momento, me dar conforto e forças para chegar onde estou.

Aos meus pais Damiana José e José Severino que foram meus maiores motivos para minha formação acadêmica, que estiveram comigo a todo momento, dedicando seu mais puro e sincero amor.

Minha gratidão a minha mãe, que durante minha jornada de longas viagens diárias até a universidade, ficava preocupada e me esperava atenta todas as noites, sempre valorizando meus esforços, faz um papel exemplar não só de mãe, mas também de pai, amiga, cúmplice, me deixou livre para seguir minhas escolhas, mas sempre me direcionando ao caminho correto, me ensinou a ser uma mulher de força, com caráter, coragem, um ser humano íntegro. Muito obrigada por sua confiança e por acreditar em mim.

Agradeço ao meu pai, que mesmo não estando muito presente, me incentivou a buscar o melhor para minha vida, e mesmo longe sempre demonstrou seu amor fraterno.

Ao meu irmão Jakson Pereira, que deu suporte quando precisei por torcer por minhas conquistas.

De forma particular, quero não apenas agradecer, mas também demonstrar meu amor incondicional e infinito a minha avó Irací Maria Pereira, que hoje se encontra ao lado de Deus, foi duas vezes mãe, primeiro porque era minha avó e segundo porque a todo momento fez papel de mãe, enquanto vida me amparou, criou e se preocupou comigo. Mesmo no céu sei que me abençoa a todo instante, é a luz que ilumina meu caminho, meu exemplo de vida. Sei que está muito feliz por minha formação acadêmica. Meu eterno agradecimento por me fazer entender que não há limites para amar.

Minha estima a todos os professores que me acompanharam desde o início da minha vida, foram incríveis na arte de ensinar, marcando assim meus maiores passos, obrigada pelo empenho, paciência e dedicação.

A todos os professores e coordenadores do curso de Estatística, que deixaram seus registros de ensinamentos durante o decorrer do curso.

Em especial reverencio o professor Tiago Almeida, por sua dedicação e orientação para com este trabalho, agradeço por me fazer enxergar a estatística com bons olhos, por

me motivar a dar continuidade a essa carreira profissional, por me aceitar como orientanda. Cada parabéns que recebi, me deu estímulo suficiente para querer o melhor sempre mais.

Quero Expressar minha satisfação aos professores participantes da banca examinadora que dividiram comigo este momento tão importante e esperado: Gustavo Henrique Esteves e Kleber Napoleão de Oliveira Barros. Obrigada pela disponibilidade e especial atenção.

Enfatizo também um agradecimento aos colegas de curso, por todos os momentos vividos durante esses anos de estudo, dividindo momentos de descontração, estudo, experiências e conquistas. Destaco Daliane Oliveira e Jonatha Lopes, pessoas que desde o primeiro dia de aula se tornaram especiais na minha vida, abriram as portas da sua casa pra mim, me apoiaram e incentivaram, e hoje agradeço por todo carinho. A Arnete Matias, que foi uma das pessoas que esteve muito próxima a mim nesses últimos anos de curso, e por ter me emprestado os livros para que este trabalho pudesse ser elaborado. E em especial agradeço a Damião Flávio e Abraão de Paula, que deram além da amizade verdadeira, a motivação e suporte para chegar até aqui, sempre torceram e me apoiaram. Obrigada pela lealdade e paciência.

Carinhosamente agradeço a todos que direta ou indiretamente fizeram parte da minha formação, garanto que este é apenas o começo de grandes vitórias.

“Hoje, neste tempo que é seu, o futuro está sendo plantado. As escolhas que você procura, os amigos que você cultiva, as leituras que você faz, os valores que você abraça, os amores que você ama, tudo será determinante para a colheita futura.”

(Padre Fábio de Melo)

Resumo

A análise de sobrevivência é utilizada em diversas áreas do conhecimento, e fundamenta-se em técnicas estatísticas que estuda dados de tempo de vida de um indivíduo. É importante que o banco de dados reúna condições necessárias pra que a análise de sobrevivência possa ser realizada e assim poder fazer estimativas cabíveis da função de sobrevivência e de risco. Este trabalho tem como objetivo utilizar conceitos do modelo semi-paramétrico de sobrevivência, em especial o modelo de Cox, para modelar o tempo de vida medido em quilômetros rodados do pneu 11.00R22. O Modelo de Cox estima a função de risco e muitas vezes é interessante estimar o efeito das covariáveis. Desta forma, Cox propôs o modelo semi-paramétrico de riscos proporcionais, que permite a análise de dados resultantes de estudos do tempo de vida, onde a variável resposta é o tempo de ocorrência de um evento de interesse ajustado por covariáveis. O estudo foi realizado com caminhões que utilizaram pneus 11.00R22, e o tempo em quilômetros rodados foi calculado através de um computador de bordo. Os dados são constituídos de um total de 552 pneus, os quais apresentaram diferença entre os três tipos de tempo de vida. Na execução deste trabalho, aplicou-se os métodos não paramétricos de Kaplan-Meier para os três grupos de vida, e o teste log-rank para verificar a existência de diferença significativa entre as curvas de sobrevivência, e em sequência, aplicamos o modelo de Cox para constatar os pressupostos de proporcionalidade. Na aplicação dos métodos foi utilizado o *software* R por meio do pacote *survival*. Os resultados encontrados para modelagem do tempo de vida dos pneus 11.00R22 indicam que o modelo de Cox estratificado é o mais adequado para modelar os dados, uma vez que foram encontradas evidências de violação da suposição de proporcionalidade.

Palavras-chaves: Modelos semi-paramétricos. Modelo de Cox. Suposição dos riscos proporcionais.

Abstract

Survival analysis is used in many areas of knowledge, and is based in statistical techniques that study data of an individual's life time. It's important that the database possesses necessary conditions for the survival analysis can be performed and so be able to make reasonable estimates of the survival function and risk. This work aims to use concepts of semi-parametric model survival, especially the Cox model to model the lifetime measured in kilometers traveled the 11.00R22 tires. The Cox model estimates the hazard function and it is interesting to estimate the effect of covariates. So Cox proposed the model semi-parametric proportional hazards, which allows the analysis of data resulting lifetime studies, where the variable answer is the time of occurrence of a event of interest adjusted for covariates. The study was carried out with trucks that used 11.00R22 tires, and the time in kilometers traveled was calculated by an onboard computer. The database consists of a total of 552 tires, which showed the difference between the three types of life time. Implementing this work, its was applied the non-parametric methods of Kaplan-Meier for the three groups, and the log-rank test to verify the existence of significant difference between the survival curves, and in sequence, we apply the Cox model to find out the assumptions proportionality. We utilized the software R with survival package in the applications. The results, indicate that the stratified Cox model is the most suitable for modeling the data, once that were found evidence of violation of the supposition proportionality.

Key-words: Semi-parametric models. Cox model. Assumption of risk Proportional.

Lista de ilustrações

Figura 1 – Representação gráfica de censura a direita	13
Figura 2 – Representação gráfica de censura a esquerda	14
Figura 3 – Representação gráfica de censura a aleatória	14
Figura 4 – Representação gráfica das estimativas de Kaplan-Meier para a função de sobrevivência para cada um dos grupos	29
Figura 5 – Suposição de riscos proporcionais fazendo uso dos resíduos padronizados de Schoenfeld	30
Figura 6 – Curvas de risco acumulado de base dos estratos obtidas por meio do modelo de Cox estratificado.	31

Lista de tabelas

Tabela 1 – Estimativas do Kaplan-Meier da função de sobrevivência para o tempo de rodagem dos pneus 11.00R22	28
Tabela 2 – Representação tabular do teste de log-rank aplicados aos grupos analisados	29
Tabela 3 – Resultado do ajuste do modelo de regressão de Cox	30
Tabela 4 – Teste de proporcionalidade das funções de risco no modelo ajustado. .	30

Sumário

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Análise de Sobrevivência	13
2.2	Processo de Contagem	16
2.3	Modelos de Regressão	16
2.3.1	Modelo de riscos proporcionais: Caso Semi-paramétrico	17
2.3.2	Modelo de tempo de vida acelerado	18
2.4	Modelo de Cox	18
2.5	Estimação no Modelo de Cox	19
2.5.1	Método da Máxima verossimilhança Parcial	20
2.6	Seleção de Modelos	22
2.6.1	Teste de Wald	22
2.6.2	Teste da Razão de Verossimilhança	22
2.7	Resíduos de Schoenfeld	23
2.8	Modelo de Cox Estratificado	24
3	MATERIAL E MÉTODOS	26
3.1	Material em Estudo	26
3.2	Métodos Estatísticos	26
4	APLICAÇÃO	28
5	CONCLUSÃO	32
5.1	Pesquisas Futuras	32
	REFERÊNCIAS	33
6	APÊNDICE A	35

1 Introdução

A análise de sobrevivência caracteriza-se em estudar o tempo de vida do indivíduo, desde o momento que ele entra no estudo, até o momento da ocorrência do evento de interesse, denominado falha. Uma particularidade é a possibilidade da presença de censura, que nada mais é, que a observação parcial da resposta de interesse, ou seja, quando o indivíduo não sofre o evento (falha) durante o período do estudo, onde o tempo exato de sobrevivência não é conhecido. É justamente nas observações censuradas que a análise de sobrevivência se diferencia. O uso de covariáveis pode afetar o tempo de vida dos indivíduos, daí surge a necessidade de fazer uso da análise de regressão.

Em meados do século XX, o uso da análise de sobrevivência foi implantado com maior frequência, e obteve maior utilidade por volta da década de oitenta, ainda no mesmo século. Mesmo podendo ser aplicado em diversas áreas do conhecimento, foi na medicina que a análise de sobrevivência chegou a sua grande popularização, porém, possui grande aplicação em diversas áreas do conhecimento humano, ou seja, não se restringe apenas a medicina. Um exemplo disso são os estudos de confiabilidade aplicados pela indústria (transportes, elétrica, etc.), onde se mede o tempo, a durabilidade de cada indivíduo em estudo.

O transporte rodoviário se intensificou no século XX, superando o transporte ferroviário no deslocamento de pessoas. E apesar de sua importância, os custos de manutenção são elevados, em comparação com os demais transportes, já que transitam com maior frequência e com cargas pesadas pelos mais diversos tipos de rodovias. Em consequência disso, a banda de rodagem (*tread*) se desgasta e fica abaixo do limite de uso recomendado, isso quer dizer que já está na hora de fazer a reforma do pneu. Daí o interesse em fazer o estudo acerca da vida útil dos pneus 11.00R22, já que o custo da produção de um pneu reformado é consideravelmente inferior ao custo de um pneu novo.

Segundo a Associação Brasileira do Segmento de Reforma de Pneus (ABR), no setor de transportes, o Brasil é o país que mais reforma pneus no mundo, devido ao seu baixo custo, reduzindo em até 57% o custo/km, podendo reformar cada pneu em média duas vezes, desta forma aumentando sua vida útil.

O presente trabalho tem por objetivo utilizar modelos semi-paramétricos de sobrevivência para modelar o tempo de vida do pneu 11.00R22, medidos em quilômetros rodados.

Neste trabalho, iremos apresentar conceitos básicos de análise de sobrevivência. Apresentaremos os modelos de regressão que dependendo da distribuição do tempo de vida dos indivíduos, pode ser paramétrico ou semiparamétrico, utilizados para modelar

a influência das covariáveis no tempo de sobrevivência. Veremos definições e suposições de violação do modelo de riscos proporcionais e qual método é o mais indicado para o estudo dos pneus 11.00R22.

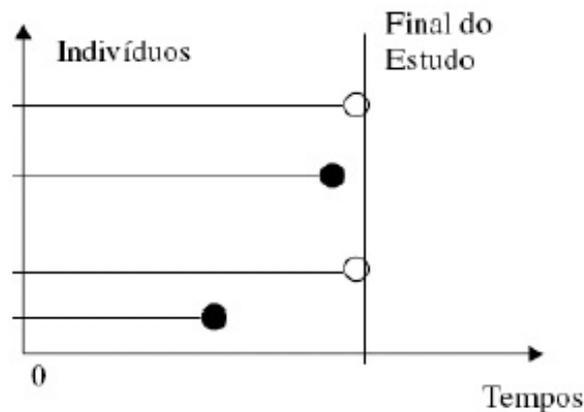
2 Fundamentação Teórica

2.1 Análise de Sobrevivência

A análise de sobrevivência é uma técnica na qual a variável resposta é determinada pelo tempo até a ocorrência do evento de interesse (falha) que ocorre depois de um tempo pré-definido (tempo de falha). Num estudo de sobrevivência, podem ocorrer observações parciais da resposta, chamada de censura. Segundo Carvalho et al. (2011) entende-se como censura, a observação incompleta do estudo por não ter observado a data de ocorrência do desfecho (falha). Podemos citar como tipos de censura:

- Censura à direita: Sendo a mais comum, pois o tempo entre o início da observação e o evento de interesse é maior que o tempo observado.

Figura 1 – Representação gráfica de censura a direita

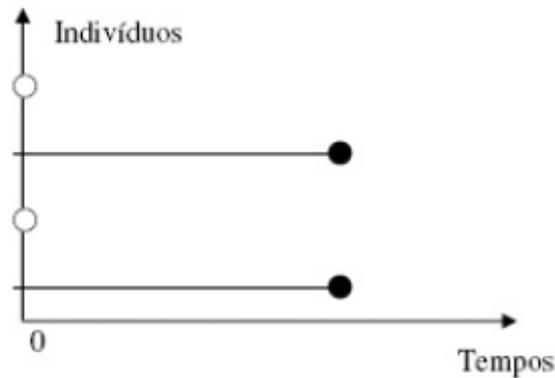


Fonte: Pascoa (2014)

Há três tipos de censura à direita, em ensaios de confiabilidade que ocorrem com mais frequência (LAWLESS, 2011), são elas: censura tipo I, existe um tempo pré-estabelecido para o acompanhamento das observações em estudo; censura tipo II, o acompanhamento das observações ocorre por um número pré-definido de falhas; censura aleatória, quando o indivíduo é retirado do estudo sem que houvesse ocorrido a falha, ou em caso de morte, em que o motivo não esteja relacionado ao estudo.

- Censura à esquerda: Quando o desfecho ocorre antes do tempo observado, e não conhecemos o momento em que o desfecho ocorreu.

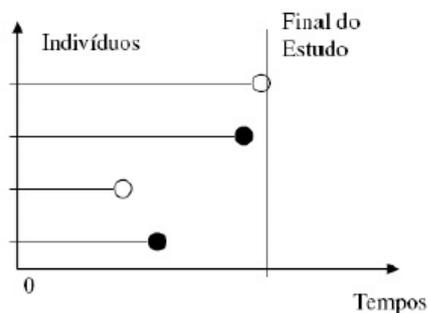
Figura 2 – Representação gráfica de censura a esquerda



Fonte: Pascoa (2014)

- Censura intervalar: O tempo de sobrevivência não necessariamente deve ser conhecido, mas sabe-se que ocorre dentro de um intervalo.

Figura 3 – Representação gráfica de censura a aleatória



Fonte: Pascoa (2014)

Sendo T uma variável aleatória não negativa, que representa o tempo de falha de um determinado indivíduo, é indicada na análise de sobrevivência o uso da função de sobrevivência ou função de taxa de falha (ou risco). A probabilidade de uma observação sobreviver até um determinado tempo t , é definida como função de sobrevivência (COLOSIMO; GIOLO, 2006), dada por:

$$S(t) = P(T > t) = 1 - F(t) = \int_0^t f(u)du, \quad t \geq 0,$$

em que $F(t)$ é a função de distribuição acumulada. Neste caso $F(t)$ representa a probabilidade de morte até o tempo t . Temos que $S(t)$ é uma função contínua e monótona decrescente, sendo $S(0) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$.

Na prática, se houver ausência de censura, a função de sobrevivência é estimada da seguinte maneira:

$$\hat{S}(t) = \frac{\text{Número de pacientes que sobrevivem após } t}{\text{Número total de pacientes}},$$

isto é, a proporção de indivíduos com o tempo de falha maior que t , portanto, uma sobrevivência empírica. Na presença de censura, $\hat{S}(t)$ não é um bom estimador para o verdadeiro $S(t)$.

A função de densidade de probabilidade é interpretada como a probabilidade de um indivíduo sofrer um evento em um intervalo instantâneo de tempo (CARVALHO et al., 2011).

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

Na prática, se não houver censura, a função $f(t)$ é estimada a partir dos tempos em que ocorrem o evento:

$$\hat{f}(t) = \frac{\text{Número de pacientes que morreram no intervalo de 0 a } t}{(\text{Número total de pacientes})(\text{Comprimento intervalo})}.$$

A função de risco $\lambda(t)$, caracteriza-se pela taxa instantânea de falha que ocorre no tempo t , dado que o indivíduo sobreviveu até o início do intervalo, e é dada por:

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

Na ausência de censura a função de risco pode ser estimada da seguinte maneira:

$$\hat{\lambda}(t) = \frac{\text{Número de falhas em cada intervalo de tempo}}{\text{Número de pessoas em risco no intervalo de tempo}}.$$

Outra função de sobrevivência que é de grande importância é a função de risco acumulado. De acordo com Carvalho et al. (2011) a função de risco acumulado $\Lambda(t)$ mede o risco de ocorrência do evento até um determinado tempo t . É definida como:

$$\Lambda(t) = \int_0^t \lambda(u) d(u) = -\log \{S(t)\}.$$

Para realizar estimação existem métodos não paramétricos e paramétricos, a citar: o estimador de Kaplan-Meier, tabela de vida e estimador de Nelson-Aalen e os paramétricos: o uso das distribuições densidade de probabilidade. Para melhor entender os métodos não paramétricos ver Klein e Moeschberger (2005), Lee e Wang (2003) e Kalbfleisch e Prentice (2011).

2.2 Processo de Contagem

Através do processo de contagem é possível provar resultados relevantes na análise de sobrevivência, permitindo acomodar censuras, truncamentos e eventos múltiplos. De maneira simples podemos compreender a análise de sobrevivência observando o par (T_i, δ_i) com $i = 1, \dots, n$, substituído por $(N_i(t), Y_i(t))$, sendo $T_i = \min(T_i^*, C_i^*)$, onde T_i^* é o tempo de falha e C_i^* representa o tempo até que ocorra a censura, e

$$\delta_i = \begin{cases} 1, & \text{se ocorrer a falha,} \\ 0, & \text{se ocorrer a censura.} \end{cases}$$

É sabido que $N_i(t)$ representa o número de eventos observados em $[0, t]$ para o indivíduo i , e

$$Y_i(t) = \begin{cases} 1, & \text{se o indivíduo } i \text{ está sob observação e em risco no instante } t; \\ 0, & \text{se o indivíduo } i \text{ não está em risco.} \end{cases}$$

Segundo Carvalho et al. (2011), define-se o processo de contagem como sendo um processo estocástico $N(t)$ com $t > 0$, em que $N(0) = 0$ e $N(t) < \infty$.

2.3 Modelos de Regressão

Em uma população homogênea é necessária a utilização das distribuições contínuas univariadas, de modo que seja possível modelar o tempo de vida. Supondo que exista heterogeneidade entre os indivíduos podendo afetar o tempo de vida, devemos então utilizar um modelo de regressão, tendo como variável resposta o tempo de vida e como variáveis independentes as covariáveis. Uma vez que se estuda os dados de sobrevivência utilizando-se uma distribuição paramétrica, pode-se ter interesse em conhecer o relacionamento entre o tempo de sobrevivência e variáveis explanatórias, neste tipo de abordagem citamos duas classes de modelos de regressão: Modelos de riscos proporcionais e o modelo de tempo de vida acelerado.

2.3.1 Modelo de riscos proporcionais: Caso Semi-paramétrico

Havendo a presença de covariáveis nos dados de sobrevivência, uma maneira de modelar esses dados é utilizando o modelo semi-paramétrico desenvolvido por Cox (1972), que é comumente utilizado em pesquisas médicas para análises de dados em tempo de sobrevivência. Este modelo descreve a relação existente entre a ocorrência de eventos. O risco é a probabilidade de um indivíduo experimentar o evento em um período mediano em torno desse ponto no tempo. O modelo de Cox ajusta a função de risco $\lambda(t)$ considerando um risco basal $\lambda_0(t)$, e o vetor de covariáveis \mathbf{x} .

Seja T , o tempo de vida de um indivíduo e num instante t , temos

$$\lambda(t|x) = \lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) = \lambda_0(t) \exp(\mathbf{x}\beta),$$

em que a função de risco $\lambda(t|x)$ é dependente de um conjunto de covariáveis (x_1, x_2, \dots, x_p) , \mathbf{x} é o vetor das variáveis explicativas, em que os coeficientes de regressão são desconhecidas e corresponde o efeito das covariáveis na sobrevivência, determinado por um vetor de componentes dos parâmetros ajustáveis β , cujo impacto é medido por seus respectivos coeficientes $(\beta_1, \beta_2, \dots, \beta_p)$.

Dizemos riscos proporcionais devido a razão entre as taxas de risco para os indivíduos com diferentes valores ou categorias das variáveis explicativas \mathbf{x} por serem ditas independentes do tempo. Também podemos escrever o modelo de riscos proporcionais em forma de função de risco acumulado ou de sobrevivência (CARVALHO et al., 2011).

$$\Lambda_0(t) = \Lambda_0(t) \exp(\mathbf{x}\beta);$$

$$S(t) = [S_0(t)]^{\exp(\mathbf{x}\beta)},$$

e a função de sobrevivência de base sendo:

$$S_o(t) = \exp\{-\Lambda_o(t)\}.$$

Breslow (1972) propôs uma estimativa acessível, formada por uma função escada com saltos nos tempos distintos de falha, com risco acumulado basal representado por:

$$\hat{\Lambda}_0(t) = \sum_{i:t_i \leq t} \frac{\Delta N_i(t)}{\sum_{j \in R(t_i)} \exp(x_j \hat{\beta})}.$$

As funções de sobrevivência $S_0(t)$ e $S(t)$ estimadas são respectivamente:

$$\hat{S}(t) = [\hat{S}_0(t)]^{\exp\{\mathbf{x} \hat{\beta}\}},$$

$$\hat{S}_0(t) = \exp(-\hat{\Lambda}_0(t)).$$

2.3.2 Modelo de tempo de vida acelerado

Esse modelo é muito utilizado quando se deseja estimar o tempo de vida de equipamentos, para isso, o teste de vida ajusta o desempenho do produto a ser testado em grandes níveis de estresse de tal forma que o produto seja obrigado a operar fora dos níveis padrões de referência, levando os resultados a apresentarem valores de tempo de vida distintos das condições normais. Para Colosimo e Giolo (2006) o tempo T obedece a seguinte expressão:

$$\ln(T) = \mu + \sigma W,$$

em que W é distribuição de probabilidade que ajusta T , μ representa a locação determinada pelo parâmetro de média $\ln(T)$ e σ também chamado de escala, representa o parâmetro de dispersão de $\ln(T)$. Há uma necessidade em utilizar esse teste para grandes amostras sob de estresse, em que o prazo a se produzir poucas ou até mesmo nenhum tipo de falha é razoável. A finalidade desse teste, como o próprio nome diz, é acelerar o evento de interesse (falha) para que se obtenha com mais rapidez as informações sobre o produto com a perspectiva de longevidade. As covariáveis agem multiplicativamente na escala do tempo, e sua função é acelerar ou retardar o tempo de vida (COLOSIMO; GIOLO, 2006). O teste de vida acelerado é comumente utilizado em indústrias eletrônicas, pois permite mostrar onde e como ocorrem falhas e usar esta informação para fazer melhores seleções dos componentes.

2.4 Modelo de Cox

Colosimo e Giolo (2006) explicam que o modelo de Cox é procedente a partir de dados que estudam o tempo de vida em que a resposta é o tempo até a ocorrência do evento de interesse, ajustado por covariáveis. Sendo o mais utilizado devido sua versatilidade, o modelo sugerido por Cox (1972), apresenta um modelo de regressão em que as observações são censuradas, ganha destaque devido a relação existente entre a função de risco e as covariáveis, e sua aplicação é possível nas mais variadas áreas, por isso a existência de um acervo significativo de artigos publicados. O modelo de Cox também pode ser utilizado para estudos relacionados a adoção de política, entre outros tipos de estudos. Para maiores informações ver (JONES; BRANTON, 2005).

Diferentemente de outros modelos, no modelo de Cox, os tempos de sobrevivência não necessitam seguir uma distribuição estatística particular, sendo uma função de risco com uma linha basal estimada não parametricamente. A razão entre os riscos de ocorrências

do evento de dois indivíduos i e j não varia ao longo do tempo, com covariáveis $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ e $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})$, temos que:

$$\frac{\lambda_i(t|x_i)}{\lambda_j(t|x_j)} = \frac{\exp(x_i\beta)}{\exp(x_j\beta)}. \quad (2.1)$$

Logo, podemos afirmar que ambas as funções descritas na expressão (2.1) para dois indivíduos quaisquer, as funções de riscos são proporcionais devido ao fato desse modelo ser considerado um modelo de riscos proporcionais, sabendo que as covariáveis apresentam na função de risco um efeito multiplicativo. Para a análise do tempo, usualmente o vetor de covariáveis não é nulo para a identificação da situação padrão. A regressão de Cox põe em prática os modelos de risco proporcionais em um ou mais preditores, e por ser semi-paramétrico, a taxa de risco de base ou a estimativa do risco absoluto não requer especificações (MARUBINI; VALSECCHI, 2004).

Tendo todos os pressupostos atendidos, o modelo de Cox possibilita melhores estimativas tanto de risco acumulado como também de probabilidades de sobrevivência, do que as da função de Kaplan-Meier. Em casos da ocorrência de variáveis contínuas, não faz sentido admitir que a covariável seja nula numa situação padrão, estipula-se assim, a média dessa covariável. Há possibilidades de manipular as covariáveis não transformadas como também a função de risco usual, tendo em vista que a modificação feita não influencia as covariáveis no risco de morte. Considere a covariável x_j , a função de risco para o i -ésimo indivíduo é descrita como:

$$h(t; x_i) = \lambda_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}).$$

2.5 Estimação no Modelo de Cox

Colosimo e Giolo (2006) conceituam que os coeficientes β 's caracterizam o modelo de Cox, este mede os efeitos das covariáveis sobre a função de taxa de falha.

Para fazer inferência quanto aos parâmetros β do modelo, é preciso fazer uso de uma técnica de estimação, ou seja, o método da máxima verossimilhança parcial, que estima o vetor de parâmetros β . Devido a presença do componente não paramétrico $\lambda_0(t)$, não é viável o método da máxima verossimilhança. Dado,

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n [f(t_i|x_i)]^{\delta_i} [S(t_i|x_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [\lambda(t_i|x_i)]^{\delta_i} S(t_i|x_i). \end{aligned} \quad (2.2)$$

Como no modelo de Cox,

$$S(t_i|x_i) = \exp \left\{ - \int_0^{t_i} \lambda_0(u) \exp \{x'_i \beta\} du \right\} = [S_0(t_i)]^{\exp \{x'_i \beta\}}.$$

Prosseguindo a aplicação deste resultado em (2.2), obtêm-se a função do componente não paramétrico $\lambda_0(t)$. Portanto temos:

$$L(\beta) = \prod_{i=1}^n [\lambda_0(t_i) \exp \{x'_i \beta\}]^{\delta_i} [S_0(t_i)]^{\exp \{x'_i \beta\}}$$

A partir daí, para que o componente não paramétrico $\lambda_0(t)$ seja eliminado, é utilizado o método da máxima verossimilhança parcial proposto por Cox (1972) que condiciona a construção da função de verossimilhança ao conhecimento prévio de falhas e censuras.

2.5.1 Método da Máxima verossimilhança Parcial

Na máxima verossimilhança, os parâmetros de um modelo são estimados, as estimativas utilizadas serão as que tornarem o valor da função de verossimilhança o *máximo* possível. De maneira inversa, também é possível estimar os parâmetros encontrando o ponto que tornar *mínima* a função log-verossimilhança negativa.

Para a máxima verossimilhança parcial Carvalho et al. (2011) explica que a mesma estima o vetor de parâmetros β , eliminando-se a função de risco basal, e apenas cada tempo t , a informação dos indivíduos sob risco é considerada. Semelhante a Kaplan-Meier, porém os efeitos das covariáveis podem ser estimados.

Considere que num estudo com uma amostra de n indivíduos observados em r tempos de vida distintos com $k \leq n$, e $t_1 < t_2 \cdots < t_r$, de maneira que $n - r$ tempos de censura a direita e t_j representa o j -ésimo tempo de vida ordenado.

Carvalho et al. (2011) explica como entender a verossimilhança parcial, considerando que cada indivíduo contribui para o tempo de sobrevivência t_i , onde as observações sob risco em t_i são conhecidas, sendo a razão entre risco do indivíduo i vir a apresentar o evento em t_i e a soma dos riscos até a ocorrência do evento em todos os indivíduos em risco,

$$L(\beta) = \frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_i)} = \frac{\exp(x_i \beta)}{\sum_{j \in R(t_i)} \exp(x_j \beta)},$$

com $\lambda_i(t_i)$ sendo o risco do indivíduo i apresentar o evento em t_i , $R(t_i)$ representa o conjunto de indivíduos em risco no instante t_i e \mathbf{x}_i como sendo o vetor de covariáveis agregados ao indivíduo que morreu no instante t_i

A inferência dos parâmetros do modelo é determinado por:

$$L(\beta) = \prod_{i=1}^n \left(\frac{\exp\{x_i\beta\}}{\sum_{j \in R(t_i)} \exp\{x_j\beta\}} \right)^{\delta_i}, \quad (2.3)$$

com δ_i sendo o indicador de falha. Apenas foram consideradas observações distintas uma vez que observações empatadas têm probabilidade nula e os tempos de sobrevivência são contínuos. Colosimo e Giolo (2006) acrescentam que em casos de ocorrer empate de falha ou censura devido a escala de medida utilizada, isto é, se utilizado uma escala discreta (horas, dias, semanas, meses ou anos) e os tempos de sobrevivência para dois ou mais indivíduos é frequente no estudo. Considera-se que a censura ocorreu após a falha, determinando portanto, quais observações devem ser incluídas no conjunto de risco em cada tempo de falha. Diante disso usa-se a aproximação proposta por Breslow (1972) e Peto (1972) para (2.3), quando o número de observações empatadas for pequeno para qualquer tempo. Seja s_i o vetor gerado pela soma das p covariáveis para os indivíduos que falharam no tempo t_i e d_i sendo o número de falhas nesses tempos com $i = 1, 2, \dots, n$, daí teremos:

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{s_i\beta\}}{\left[\sum_{j \in R(t_i)} \exp\{x_j\beta\} \right]^{d_i}},$$

em caso contrário o indicado é utilizar o modelo de Cox para dados agrupados (LAWLESS, 1982). Utilizando a função de máxima verossimilhança, a inferência acerca dos parâmetros do modelo de Cox, é feita aplicando as estatísticas de Wald, Razão de verossimilhança e Escore. A função $L(\beta)$ considerada por Cox, não é uma verossimilhança usual, em casos de censura não informativa,

$$L(\beta) = \prod_{i=1}^n [h(t_i)]^{\delta_i} S(t_i),$$

que para o modelo de Cox tomaria a forma:

$$\begin{aligned} L[\beta, \lambda_0(t)] &= \prod_{i=1}^n \left[\lambda_0(t_i) \exp(\beta' z_i) S_0(t_i)^{\exp(\beta' z_i)} \right]^{\delta_i} \left[S_0(t_i)^{\exp(\beta' z_i)} \right]^{1-\delta_i} \\ &= \prod_{i \in D} \frac{\exp(\beta' z_i)}{\sum_{l \in R_i} \exp(\beta' z_l)} \prod_{i \in D} (h_0(t_i) \sum_{l \in R_i} \exp(\beta' z_l)) \prod_{i=1}^n S_0(t_i)^{\exp(\beta' z_i)}, \end{aligned}$$

sendo D o conjunto de indivíduos da qual a morte foi observada. A partir do momento que $L(\beta)$ corresponde com o primeiro fator de $L[\beta, h_0(t)]$, representa uma verossimilhança parcial. Não há necessidade de especificar $\lambda_0(t)$ para que haja inferência sobre o vetor de parâmetros β tendo em vista que $L(\beta)$ não depende de $h_0(t)$. Sob situações de regularidade

gerais, Andersen e Gill (1982) determinaram que o estimador de máxima verossimilhança parcial β são consistentes e assintoticamente normais, com matriz de covariâncias $I(\beta)^{-1}$,

$$I_{jk}(\beta) = E \left(\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k} \right).$$

2.6 Seleção de Modelos

Utilizam-se os mesmos métodos aplicados aos modelos lineares generalizados, podendo assim fazer comparações aos diferentes modelos de regressão de Cox. Os testes utilizados são o Teste de Wald e o teste da razão de probabilidade ou de verossimilhança. O conjunto de testes baseados na verossimilhança, são dados pela hipótese nula simples $H_0 : \beta = \beta_0$. Estas estatísticas podem ser usadas para testar outras hipóteses. Quando o modelo de Cox não se ajusta a uma situação clínica, o ideal é utilizar técnicas que avaliem sua adequação. Deste modo, a violação da suposição básica, que é a de taxas de falha proporcionais, pode acarretar em sérios vícios na estimação dos coeficientes do modelo (STRUTHERS; KALBFLEISCH, 1986). Podem utilizar coeficientes que variam com o tempo, ou melhor, como o risco se modificam com o tempo, algumas covariáveis podem não presumir os riscos proporcionais. Assim sendo, o modelo de Cox estratificado é um forte candidato, quando a suposição de riscos proporcionais não estiver satisfeita dentro de cada estrato.

2.6.1 Teste de Wald

O teste de Wald é fundamentado em amostras grandes dos estimadores de máxima verossimilhança $\hat{\beta}$ que têm distribuição normal multivariada, e é definida por:

$$\chi_W^2 = (\hat{\beta} - \beta_0)^t I(\hat{\beta})(\hat{\beta} - \beta_0).$$

Para quando H_0 é verdadeira, com distribuição qui-quadrado e p graus de liberdade para amostras grandes.

2.6.2 Teste da Razão de Verossimilhança

O teste da razão de probabilidades é baseada na estatística assintótica qui-quadrada com p graus de liberdade sob hipótese nula. Definida pela diferença entre os logaritmos da função de verossimilhança:

$$\chi_{LR}^2 = 2[\ln L(\beta_0) - \ln L(\hat{\beta})].$$

De acordo com Carvalho et al. (2011) o teste da razão de verossimilhança faz uma comparação nos modelos aninhados, verificando se, quando houver inclusão de uma ou

mais covariáveis no modelo, a verossimilhança de um modelo aumenta significativamente, em relação ao modelo parcimonioso. Os autores ainda complementam que a razão de verossimilhança é assintoticamente semelhante à estatística de Wald quando o número de observações for grande.

2.7 Resíduos de Schoenfeld

Os resíduos são definidos apenas para falhas e são importantes para testar se o modelo de regressão é adequado ou não. Os resíduos de Schoenfeld (1982) averiguam a proporcionalidade de cada covariável z_i , definidos para cada i -ésimo indivíduo como:

$$r_{ji} = \delta_i(z_{ji} - a_{ji}),$$

em que;

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é não censurado,} \\ 0, & \text{se } t_i \text{ é censurado.} \end{cases}$$

Para esse indivíduo, tem-se um vetor de resíduos $\mathbf{r}_j = (r_{j1}, r_{j2}, \dots, r_{jp})$ e i ($i = 1, \dots, p$), indicando os tempos observados de eventos, R_i sendo o conjunto dos indivíduos em risco no instante t e z_{jl} o valor da covariável l do indivíduo j , dado por:

$$a_{ji} = \frac{\sum_{l \in R_i} z_{jl} \exp(\hat{\beta}' z_l)}{\sum_{l \in R_i} \exp(\hat{\beta}' z_l)}.$$

Por definição, quando uma observação é censurada, o resíduo tem o valor zero. De acordo com Colosimo e Giolo (2006), o conjunto de resíduos de Schoenfeld é uma matriz com d número de falhas que representam as linhas e p covariáveis consideradas no modelo, representando portanto as colunas.

A verossimilhança parcial $L(\beta)$ verifica a igualdade:

$$\frac{\partial \log L(\hat{\beta})}{\partial \hat{\beta}_j} = \sum_{i=1}^n r_{ji} = 0,$$

em que $\hat{\beta}'$ com $j = 1, \dots, p$, são as estimativas de máxima verossimilhança dos coeficientes β_j das covariáveis, desta forma, para cada covariável, a soma dos resíduos de todos os indivíduos em estudo é sempre nula.

Segundo Carvalho et al. (2011), supondo que um coeficiente β varia com o tempo t , com β podendo ser dividido em duas partes: uma média constante com variância $V(\beta)$ e

uma função $U(t)$ que varia no tempo. Desta forma, os resíduos padronizados de Schoenfeld em t_i são representados por:

$$r_i^* = \frac{r_i(\beta)}{V(\beta)}.$$

O valor esperado de r_i^* para cada grupo em $R(t_i)$ é aproximadamente igual a parte de β que varia no tempo. Grambsch e Therneau (1994) apresentam uma versão padronizada destes resíduos para avaliar proporcionalidade dos riscos após os ajustes do modelo de Cox:

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

restringindo $\beta(t) = \beta$, acarretando na proporcionalidade dos riscos.

Para análise do gráfico dos resíduos de Schoenfeld *versus* o tempo de vida, verifica-se se os dados estão distribuídos igualmente ao longo do tempo e dispostos numa nuvem de pontos aleatórios, centrados em zero, numa linha horizontal. Sendo assim, a suposição de riscos proporcionais permite um modelo adequado para os dados, não havendo tendências sistemáticas no gráfico r_i^* contra o tempo de sobrevivência.

Há possíveis evidências de proporcionalidade quando houver inclinação zero, uma curva suavizada é adicionada com bandas de confiança, facilitando a identificação de possíveis falhas da suposição de riscos proporcionais (COLOSIMO; GIOLO, 2006).

Analisando o valor p , se próximo a zero, configura-se possivelmente que não há rejeição da suposição dos riscos proporcionais. Em caso contrário, existem evidências sugestivas de não proporcionalidade, neste caso, possíveis soluções são acionadas, como por exemplo, particionar o eixo do tempo, estratificação pela covariável tempo-dependente, ou o uso de outro tipo de modelo.

2.8 Modelo de Cox Estratificado

Utilizado quando a suposição de riscos proporcionais for violada em todos os indivíduos. Em casos de não haver evidências que o risco basal $\lambda_0(t)$ seja igual para todos os indivíduos, a separação em diferentes grupos com $\lambda_{0A}(t) \neq \lambda_{0B}(t) \neq \lambda_{0C}(t)$, que define diferentes estratos (CARVALHO et al., 2011). Vale ressaltar que o efeito da variável para a qual se estratifica o risco basal, não será estimado.

O modelo de Cox estratificado, fundamenta-se de acordo com a indicação de violação da suposição, que consiste em fragmentar os dados de sobrevivência em m

estratos, representado por:

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp(X'_{ij}\beta),$$

dado que $i=1, \dots, n_j$ e $j= 1, \dots, m$, com n_j representando o número de observações no j -ésimo estrato.

O vetor β é comum em todos os estratos e o risco basal não se relacionam e varia para cada estrato. Calcula-se a verossimilhança individual da mesma maneira que a verossimilhança parcial do modelo de Cox, restringindo apenas os indivíduos que pertencem ao mesmo estrato. A verossimilhança parcial total para o risco é baseado no produto das verossimilhanças parciais para cada estrato (CARVALHO et al., 2011), ou seja:

$$L(\beta) = \prod_{j=1}^s L_j(\beta).$$

Carvalho et al. (2011), acrescenta que, no modelo de Cox estratificado, os efeitos estão controlados pela variação no risco basal atribuível aos estratos. Caso o modelo de Cox estratificado seja usado sem que haja necessidade, acarretará na eliminação da eficácia das estimativas obtidas.

3 Material e Métodos

Nesta seção serão apresentados os dados e os métodos estatísticos para o desenvolvimento neste trabalho.

3.1 Material em Estudo

O material utilizado para elaborar o estudo, foi baseado no banco de dados utilizado no trabalho de Santos (2016), cuja população em estudo, foram 552 pneus do tipo 11.00R22 de caminhões transportadores de safras, afim de compreender a aplicação do modelo de Cox quando há violação na suposição dos riscos, onde foi feito uma revisão teórica com suas principais definições.

Para os dados em estudo, o tempo de vida do pneu foi medido em quilômetros rodados, com interesse de avaliar o tempo de cada vida útil, já que o pneu passava por um processo de reforma quando ocorria a falha ou censura, podendo ter até três vidas úteis.

3.2 Métodos Estatísticos

Utilizou-se modelos básicos de sobrevivência, o estimador de Kaplan-Meier e o teste de log-rank, com propósito de verificar se existiam diferenças estatísticas do tempo de vida por grupos. Foram utilizados modelos de Cox de riscos proporcionais e para verificar indícios de violação utilizou-se os resíduos schoenfeld, e por fim foi aplicado o modelo de Cox estratificado. O procedimento realizado para analisar a sobrevivência de um total de 552 pneus 11.00R22 foi o seguinte:

1. Construção do gráfico de Kaplan-meier, permitindo observar o comportamento dos dados ao longo do tempo em relação a sobrevivência.
2. Aplicação do teste de log-rank para verificar a possível existência de diferença significativa entre os grupos de vida 1, 2 e 3.
3. Verificação os pressupostos de proporcionalidade de cada grupo através do ajuste do modelo de Cox e dos resíduos de Schoenfeld.
4. Estratificação da função de risco por meio do modelo de Cox estratificado, devido a violação dos riscos proporcionais.

Todas as análises aqui apresentadas, foram feitas com o auxílio do *software* R versão 3.3.1 (R Core Team, 2014) através do pacote *survival* versão 2.39-5 (THERNEAU; LUMLEY, 2016).

4 Aplicação

A base de dados é composta por 552 pneus 11.00R22, que foram classificados em três grupos (Vida 1, Vida 2, Vida 3). Baseado no trabalho de Santos (2016), aplicou-se o método de Kaplan-Meier e o teste de log-rank, para estimar a função de sobrevivência. A seguir serão apresentados os principais resultados das análises realizadas com os dados dos pneus 11.00R22. A Tabela 1 apresenta as estatísticas descritivas estimadas para cada um dos grupos através do estimador de Kaplan-Meier.

Tabela 1 – Estimativas do Kaplan-Meier da função de sobrevivência para o tempo de rodagem dos pneus 11.00R22

Grupo	n	Evento/falha	Tempo mediano
Vida 1	552	449	25953
Vida 2	416	195	36781
Vida 3	138	5	NA
Total	1106	649	32229

De acordo com a Tabela 1 nota-se que o grupo 1 (Vida 1) possui 552 pneus dos quais 449 falharam durante o estudo, o mesmo apresentou um tempo mediano de 25953 km. No grupo 2 (Vida 2) é composto por 416 pneus sendo que 195 deles sofreram o evento de interesse durante o estudo, com um tempo mediano de 36781, e por fim o grupo 3 (Vida 3) é formado por 138 pneus que estavam aptos, e 5 deles experimentaram falha, o *software* não calculou a estimativa do tempo mediano do grupo 3. Totalizando assim, em um tempo mediano das três vidas úteis em aproximadamente 32000 km, na qual, 649 (58,67 %) dos pneus avaliados apresentaram alguma falha em algum período. As curvas de sobrevivência estimadas por meio do estimador de Kaplan-Meier estão mostradas para as três vidas na Figura 4.

conseguimos ver através da função escada o comportamento dos dados ao longo do tempo em relação a sobrevivência. Podemos observar na Figura 4 que, aos 10.000 km não houve diferença entre os três grupos estudados, diferentemente do que acontece após, através do prolongamento do tempo de vida de cada grupo. Aos 30.000 km, 50 % dos indivíduos já falharam. O grupo vida 1, aos 20.000 km apresenta menor valor de sobrevivência em relação aos demais grupos. O grupo vida 2 obteve maior tempo de vida mediano, e o grupo vida 3 apresentou maior sobrevivência, porém tem menor quantidade de pneus, não se aproxima de zero e não há evidências de falha aos 50.000 km, seguindo constante após os 20.000 km.

A estatística de log-rank mostra que sob hipótese de igualdade das curvas de

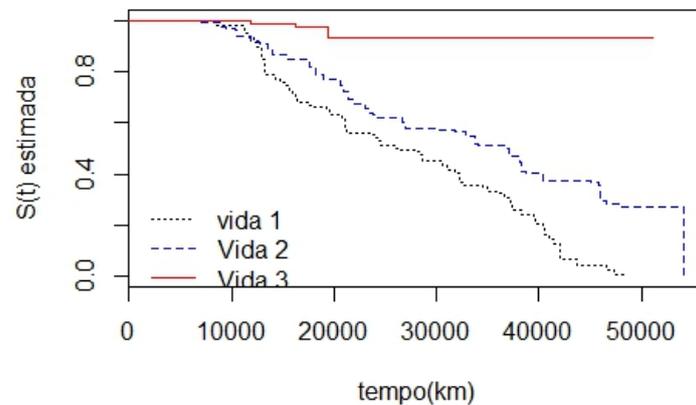


Figura 4 – Representação gráfica das estimativas de Kaplan-Meier para a função de sobrevivência para cada um dos grupos

sobrevivência, tem-se distribuição qui-quadrado com 2 graus de liberdade e resultou em $\chi^2=133$ gerando um p -valor $\cong 0$.

Em conformidade com a Tabela 2, foi constatado a existência de diferença significativa entre os grupos 1, 2 e 3, ou seja, há uma diferença no tempo de vida entre os pneus desses grupos. Pneus do grupo vida 1, apresentam tempo médio de falha menor que os demais grupos. Sendo assim, o teste de log-rank apenas confirma o que foi verificado o gráfico da Figura 4.

Tabela 2 – Representação tabular do teste de log-rank aplicados aos grupos analisados

Log-Rank	N	Informação	Valor numérico
grupo 1	552	Qui-quadrado	73,84
grupo 2	416	P-valor	0
Log-Rank	N	Informação	Valor numérico
grupo 2	416	Qui-quadrado	34,79
grupo 3	138	P-valor	$3,66 \times 10^{-09}$
Log-Rank	N	Informação	Valor numérico
grupo 1	552	Qui-quadrado	72,73
grupo 3	138	P-valor	0

Verifica-se que *Vida* é uma variável significante, indicando que ela contribui para a explicação da função de risco. A Tabela 3, afirma que uma vez ajustado o modelo de Cox, deve-se verificar se os pressupostos de proporcionalidade ao longo do tempo são razoáveis. Para isso, usaremos os resíduos padronizados de Schoenfeld, que verifica de forma mais precisa a proporcionalidade da função de risco.

A análise para o modelo de Cox foi realizado sem iteração, mostrados na Tabela

Tabela 3 – Resultado do ajuste do modelo de regressão de Cox

Covariável	Estimativas	Erro padrão	Valor p	RR
Vida 2	-0,732	0,088	2e-16	0,480
Vida 3	-2,794	0,449	5,28e-10	0,061

4, de acordo com o p -valor encontrado ($p < 0,05$), a um nível de 5 % de significância, há evidências de violação em todas as variáveis, como o p -valor global é menor que 0,05, o fator vida não foi proporcional, ou seja, é uma variável tempo-dependente.

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

Covariáveis	$\rho(\rho)$	χ^2	P-valor
Vida 2	-0,1754	21,02	4,55e-06
Vida 3	-0,0492	1,56	2,11e-01
Global	NA	22,00	1,67e-05

Para confirmação, pode-se perceber a partir da Figura 5, que os resíduos não tem um padrão aleatório em torno de zero, os resíduos situados aproximadamente a 38.000 km, apresentam um padrão decrescente, sugerindo violação do princípio de proporcionalidade da função de risco.

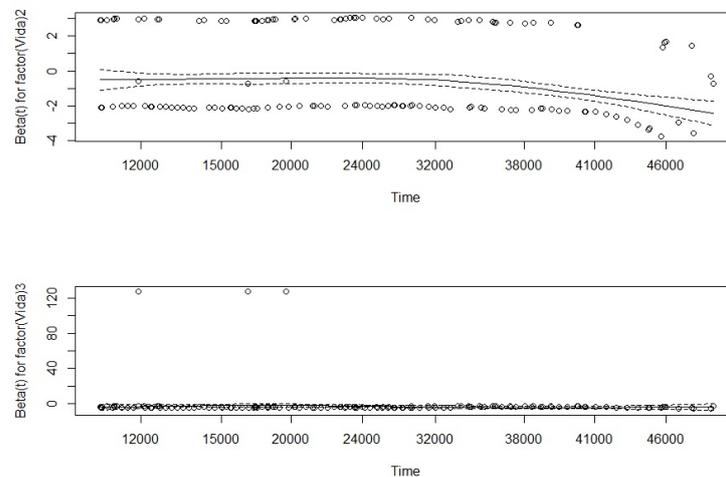


Figura 5 – Suposição de riscos proporcionais fazendo uso dos resíduos padronizados de Schoenfeld

Após a verificação da não proporcionalidade, procedeu-se para a estratificação da função de risco por meio do modelo de Cox estratificado. Como só dispomos de uma covariável e esta é a covariável de estratificação, o modelo de Cox estratificado será apresentado apenas através da função de base em cada grupo. A Figura 6 mostra as curvas obtidas pelo ajuste da função de risco via modelo de Cox.

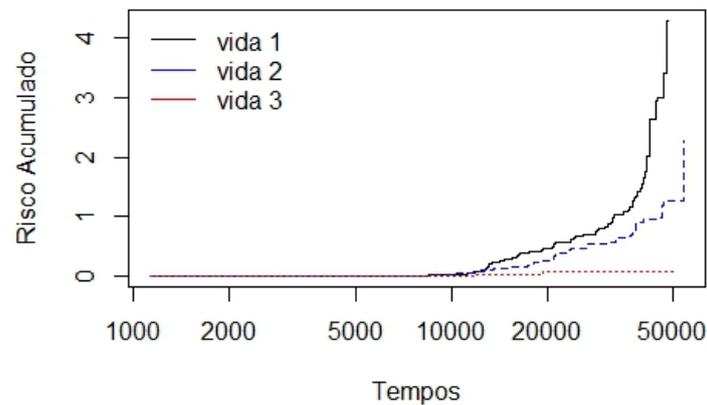


Figura 6 – Curvas de risco acumulado de base dos estratos obtidas por meio do modelo de Cox estratificado.

Cada linha representa um fator, que por sua vez representa um estrato. A partir de 20.000 km rodados, nota-se um crescimento no risco acumulado, onde os estratos com vida inicial (vida 1), evidencia risco de falha maior. Esse aumento após um pequeno intervalo de tempo inicial tende a aparecer constante. As funções de risco acumulado de base obtidas para o estrato se cruzam, reforçando pois, o indício de violação da suposição de riscos proporcionais. Os resultados mostram que o modelo de Cox estratificado para a análise dos pneus 11.00R22 parece ser o mais indicado, já que os riscos não apresentaram proporcionalidade entre os estratos.

Em estudo de confiabilidade de pneus 295/80R22.5 sucateados de frotas de transporte rodoviário de carga apresentado por Haviaras (2005), verifica-se que as falhas com maiores percentuais de perda são responsáveis pela redução da vida média dos pneus, isto é caracterizado por taxa de falhas crescentes que indicam o fim da vida útil dos pneus.

5 Conclusão

O estudo realizado com pneus 11.00R22, teve como objetivo fazer uso do modelo de Cox tradicional e Cox estratificado, devido ao fato de haver violação na suposição dos riscos proporcionais. Foram aplicados os métodos básicos e introdutórios de análise de sobrevivência.

Após a aplicação das estatísticas descritivas, observou-se o comportamento das covariáveis em estudo. A tabela de frequência constatou que o tempo de vida mediano para a troca dos pneus foi de aproximadamente 32000 Km, para um conjunto com 552 pneus.

A curva de Kaplan-Meier foi ajustada aos dados, indicando a existência de diferença entre os grupos, para comprovação, o teste log-rank confirmou haver realmente a existência significativa entre os três grupos.

No modelo de Cox de riscos proporcionais, a variável grupo foi significativa ao nível de 5% de significância. Após o modelo de Cox ter sido ajustado, fez-se a análise dos pressupostos de proporcionalidade ao longo do tempo e os resíduos de Schoenfeld evidenciaram a existência de violação desta em todos os grupos.

Em contrapartida, o modelo de Cox estratificado foi utilizado e os resultados reforçaram a evidência de violação da suposição dos riscos proporcionais. O ajuste do modelo estratificado no período de uso dos pneus, mostrou que o fator *Vida* influencia no tempo até que o pneu atinja a quilometragem máxima. Concluímos que o modelo de Cox estratificado é o mais indicado que o modelo de Cox para modelagem no tempo de vida dos pneus 11.00R22.

5.1 Pesquisas Futuras

Tendo em vista os resultados obtidos neste trabalho, é possível com a obtenção de outras covariáveis referentes a este experimento, analisar modelos de regressão dentro do modelo de cox estratificado e assim usar a razão de riscos.

Referências

- ANDERSEN, P. K.; GILL, R. D. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, JSTOR, p. 1100–1120, 1982. Citado na página 22.
- BRESLOW, N. Contribuição à discussão do artigo de dr cox. *Journal of the Royal Statistical Society, Series B*, v. 34, p. 216–7, 1972. Citado 2 vezes nas páginas 17 e 21.
- CARVALHO, M. S. et al. *Análise de Sobrevivência: teoria e aplicações em saúde*. [S.l.]: SciELO-Editora FIOCRUZ, 2011. Citado 9 vezes nas páginas 13, 15, 16, 17, 20, 22, 23, 24 e 25.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Edgard Blücher, 2006. Citado 6 vezes nas páginas 14, 18, 19, 21, 23 e 24.
- COX, P. R. *Life Tables*. [S.l.]: Wiley Online Library, 1972. Citado 3 vezes nas páginas 17, 18 e 20.
- GRAMBSCH, P. M.; THERNEAU, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, Biometrika Trust, v. 81, n. 3, p. 515–526, 1994. Citado na página 24.
- HAVIARAS, G. J. *Metodologia para análise de confiabilidade de pneus radiais em frotas de caminhões de longa distância*. Tese (Doutorado) — Universidade de São Paulo, 2005. Citado na página 31.
- JONES, B. S.; BRANTON, R. P. Beyond logit and probit: Cox duration models of single, repeating, and competing events for state policy adoption. *State Politics & Policy Quarterly*, SAGE Publications, v. 5, n. 4, p. 420–443, 2005. Citado na página 18.
- KALBFLEISCH, J. D.; PRENTICE, R. L. *The statistical analysis of failure time data*. [S.l.]: John Wiley & Sons, 2011. v. 360. Citado na página 16.
- KLEIN, J. P.; MOESCHBERGER, M. L. *Survival analysis: techniques for censored and truncated data*. [S.l.]: Springer Science & Business Media, 2005. Citado na página 16.
- LAWLESS, J. *Statistical methods and model for lifetime data*. *Wiley& Sons, New York*, 1982. Citado na página 21.
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. [S.l.]: John Wiley & Sons, 2011. v. 362. Citado na página 13.
- LEE, E. T.; WANG, J. *Statistical methods for survival data analysis*. [S.l.]: John Wiley & Sons, 2003. v. 476. Citado na página 16.
- MARUBINI, E.; VALSECCHI, M. G. *Analysing survival data from clinical trials and observational studies*. [S.l.]: John Wiley & Sons, 2004. v. 15. Citado na página 19.
- PASCOA, M. A. R. d. Intervalos de credibilidade para a razão de riscos do modelo de cox, considerando estimativas pontuais bootstrap. UNIVERSIDADE FEDERAL DE LAVRAS, 2014. Citado 2 vezes nas páginas 13 e 14.

- PETO, R. Contribuição à discussão do artigo de dr cox. *Journal of the Real Statistics Society B*, v. 34, p. 205–207, 1972. Citado na página 21.
- R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. [S.l.]: ISBN 3-900051-07-0, 2014. Citado na página 27.
- SANTOS, I. P. Introdução à análise de confiabilidade: uma aplicação ao setor de transportes. 2016. Citado 2 vezes nas páginas 26 e 28.
- SCHOENFELD, D. Partial residuals for the proportional hazards regression model. *Biometrika*, Biometrika Trust, v. 69, n. 1, p. 239–241, 1982. Citado na página 23.
- STRUTHERS, C. A.; KALBFLEISCH, J. D. Misspecified proportional hazard models. *Biometrika*, Biometrika Trust, v. 73, n. 2, p. 363–369, 1986. Citado na página 22.
- THERNEAU, T. M.; LUMLEY, T. *Package ‘survival’*. [S.l.]: Verze, 2016. Citado na página 27.

6 APÊNDICE A

**Programa em linguagem R utilizado
para análise**

```
# # Estimativas de Sobrevida
ekm=survfit(Surv(PeriodoUso,Censura) factor(Vida))
ekm
par(mfrow=c(1,1))
# # # Gráfico de Kaplan-Meier
plot(ekm, lty=c(3,2,1), col=c("black","blue", "red"),
xlab="tempo(km)",ylab="S(t) estimada")
legend(1,0.34,lty=3:1,col=c("black","blue", "red"),
c("vida 1","Vida 2","Vida 3"),lwd=1,bty="n")
survdif(Surv(PeriodoUso,Censura) factor(Vida),rho=0)
# # # Modelo de Cox
# # # fit0 = intercepto e fit1 = com covriável vida
fit0=coxph(Surv(PeriodoUso,Censura) 1,data=analise,x=T,method="breslow")
summary(fit0)
fit0$loglik
fit1=coxph(Surv(PeriodoUso,Censura) factor(Vida),data=analise,
x=T,method="breslow")
summary(fit1)
fit1$loglik
anova(fit1,fit0, test = "Chisq")
# # # modelo de Cox sem iteração
mresid <- resid(fit1, type="scaledsch")
cox.zph(fit1,transform="identity")
par(mfrow=c(2,1))
plot(cox.zph(fit1))
fit2 = com estratificação (violação de proporcionalidade)
fit2=coxph(Surv(PeriodoUso,Censura) strata(Vida),data=analise,
x=T,method="breslow")
summary(fit2)
fit2$loglik
```

```
anova(fit2,fit1)

# # # modelos separados por Vidas
res.separate <- lapply(split(analise, analise$Vida),
FUN = function(DF) coxph(Surv(PeriodoUso,Censura) 1, DF))
res.separate

# # # Plot por vidas
plot(survfit(fit2),
log = "x", fun = function(S) -log(S),
lty = rep(1:3, 1), col = rep(c("black","blue","red"), each =1),
xlab = "Tempos", ylab = "Risco Acumulado")
legend("topleft",legend=c("vida 1", "vida 2","vida 3"),
lty=1,col = c("black","blue","red"),lwd=1, bty="n")
```