

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE ELETROTÉCNICA
ESPECIALIZAÇÃO EM ENGENHARIA DA CONFIABILIDADE**

DIOGO LUIZ MARTIN

**MODELO DE RISCOS PROPORCIONAIS: UMA APLICAÇÃO AOS
DADOS DE RECLAMAÇÕES DE CONSUMIDORES**

MONOGRAFIA DE ESPECIALIZAÇÃO

CURITIBA

2017

DIOGO LUIZ MARTIN

**MODELO DE RISCOS PROPORCIONAIS: UMA APLICAÇÃO AOS
DADOS DE RECLAMAÇÕES DE CONSUMIDORES**

Monografia apresentada como requisito parcial à obtenção do título de Especialista em Engenharia da Confiabilidade, do Departamento Acadêmico de Eletrotécnica, da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Emerson Rigoni

CURITIBA

2017



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Curitiba
Diretoria de Pesquisa e Pós-Graduação
Departamento Acadêmico de Eletrotécnica
Especialização em Engenharia da Confiabilidade



TERMO DE APROVAÇÃO

MODELO DE RISCOS PROPORCIONAIS: UMA APLICAÇÃO AOS DADOS DE RECLAMAÇÕES DE CONSUMIDORES

por

DIOGO LUIZ MARTIN

Esta monografia foi apresentada em 03 de novembro de 2017, como requisito parcial para obtenção do título de Especialista em Engenharia da Confiabilidade, outorgado pela Universidade Tecnológica Federal do Paraná. O aluno foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Prof. Emerson Rigoni, Dr. Eng.
Professor Orientador - UTFPR

Prof. Carlos Henrique Mariano Dr.
Membro Titular da Banca - UTFPR

Prof. Marcelo Rodrigues Dr.
Membro Titular da Banca - UTFPR

O Termo de Aprovação assinado encontra-se na Coordenação do Curso.

Dedico este trabalho:

Ao Deus do meu coração e da minha compreensão por sua generosidade e pela oportunidade de evolução.

À minha mãe que, com muito carinho e apoio, se esforçou sobremaneira para me criar e educar.

À minha amada esposa Christianna, com quem escolhi partilhar minha vida, pelo carinho, paciência e amor.

Aos meus filhos, Mauricio e Arthur, que iluminam e motivam cada dia, hora e minuto da minha existência.

AGRADECIMENTOS

Este trabalho é a manifestação formal de uma inquietude intelectual e do empenho em ampliar os meus conhecimentos. Nesta ampliação a ajuda daqueles que contribuíram para a realização deste desejo foi fundamental. A estes, o meu profundo agradecimento:

Ao Deus do meu coração e da minha compreensão pelas dádivas de saúde, sabedoria e coragem que me permitem seguir em frente todos os dias.

À minha mãe, uma heroína, que apesar de todas as dificuldades criou-me e suportou-me durante meus primeiros passos.

À minha esposa Christianna pelo amor, incentivo e apoio incondicional.

A todos os professores da minha graduação, em especial à professora Suely Ruiz Giolo e aos professores Adilson dos Anjos e Anselmo Chaves Neto, que compartilharam os seus conhecimentos estatísticos e não somente ensinaram-me, mas me fizeram aprender.

Não deixe para a tarde o que pode fazer pela manhã. (Autor desconhecido)

Eu nunca perco. Ou ganho ou aprendo. (MANDELA, Nelson, 1995)

RESUMO

MARTIN, Diogo Luiz. **Modelo de Riscos Proporcionais: uma aplicação aos dados de reclamações de Consumidores**. 2017. 46. Monografia (Especialização em Engenharia da Confiabilidade) - Universidade Tecnológica Federal do Paraná. Curitiba, 2017.

O emprego de técnicas estatísticas, em especial a análise de sobrevivência, para análise de dados cresceu substancialmente nas últimas décadas. O foco deste trabalho é o modelo de risco proporcionais no qual o estudo do tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse é ajustado por covariáveis. Dentro dos modelos semi-paramétricos, o modelo de Cox apresenta uma flexibilidade interessante e é muito utilizado em dados de sobrevivência nos quais o efeito das covariáveis de interesse pode variar ao longo do tempo de duração do estudo. Desta maneira, o objetivo desse trabalho é testar e avaliar a capacidade do modelo de Cox de modelar com precisão ou de forma satisfatória o tempo de solução de conflitos utilizando uma ferramenta por meio da internet que permite a interlocução direta entre consumidores e empresas. Durante o trabalho, algumas dificuldades se apresentaram e foram solucionadas ou contornadas. Observou-se que algumas regiões, o sexo e algumas faixas etárias sensibilizam o entendimento do consumidor em termos de resolução das reclamações feitas no portal Consumidor.com.br.

Palavras-chave: Análise de Sobrevivência. Modelo de Riscos Proporcionais. Reclamações de consumidores. Aplicação.

ABSTRACT

MARTIN, Diogo Luiz. **Proportional Hazard Model: an application to consumer complaints data**. 2017. 46. Monografia (Post-Graduation in Reliability Engineering) - Universidade Tecnológica Federal do Paraná. Curitiba, 2017.

The usage of Statistical techniques, especially survival analysis, for data analysis has grown substantially in recent decades. This paper focus on the proportional risk model in which the study of the life time in which the response is the time until the occurrence of an event of interest is adjusted by covariables. Within the semi-parametric models, the Cox model presents interesting flexibility and is widely used in survival data in which the effect of covariates of interest may vary over the duration of the study. In that sense, the objective here is to test and to evaluate the ability of the Cox model to accurately or satisfactorily model the time of conflict resolution using a tool through the Internet that allows direct interaction between consumers and companies. During the job, some difficulties have been faced and were resolved or circumvented. In the end, some regions, sex and some age groups were identified as relevants regarding the resolution of the complaints made in the portal Consumidor.com.br.

Key words: Survival Analysis. Proportional Hazard Model. Customer Complaints. Application.

LISTA DE ILUSTRAÇÕES

Figura 2.1 – Distribuição das reclamações por Região	21
Figura 2.2 – Reclamações por Sexo	Erro! Indicador não definido. 22
Figura 2.3 – Faixa Etária dos Reclamantes.....	Erro! Indicador não definido. 2
Figura 2.4 – Situação e Avaliação das Reclamações	23
Figura 2.5 – Situação e Avaliação das Reclamações Cruzadas	23
Figura 2.6 – Tempo de Resposta	24
Figura 4.1 – Curva de Kaplan-Meier para covariável Região	3 Erro! Indicador não definido.
Figura 4.2 – Curva de Kaplan-Meier para covariável Sexo	Erro! Indicador não definido.
Figura 4.3 – Curva de Kaplan-Meier para covariável Faixa Etária	Erro! Indicador não definido.
Figura 4.4 – Suposição de riscos proporcionais para as covariáveis utilizando gráficos do logaritmo da função de risco acumulada versus tempo para as covariáveis	37
Figura 4.5 – Gráfico dos resíduos de Schoenfeld para cada variável .	Erro! Indicador não definido.
Figura 4.6 – Curvas das Estimativas da Função de Sobrevida ($S_0(t)$) ajustadas para a covariável Região.....	41
Figura 4.7 – Curvas das Estimativas da Função de Sobrevida ($S_0(t)$) ajustadas para a covariável Sexo	Erro! Indicador não definido.
Figura 4.8 – Curvas das Estimativas da Função de Sobrevida ($S_0(t)$) ajustadas para a covariável Faixa Etária	42
Quadro 4.1 – Variáveis Dummies para cada classe.....	30
Quadro 4.2 – Modelo Final Antes da avaliação da suposição de riscos proporcionais	36
Quadro 4.3 – Resumo dos resultados.....	43

LISTA DE TABELAS

Tabela 2.1 – Variáveis coletadas das reclamações	Erro! Indicador não definido.
Tabela 4.1 – Número de Reclamações com Avaliação Recebida vs Resposta da Empresa	Erro! Indicador não definido. 8
Tabela 4.2 – Número de Reclamações com Avaliação Recebida vs Resposta da Empresa já considerando as exclusões supramencionadas	Erro! Indicador não definido. 9
Tabela 4.3 – RG_V1 vs Covariável Região	30
Tabela 4.4 – Passo 1 - Seleção de covariáveis usando o modelo de regressão de COX.....	33
Tabela 4.5 – Passo 2 - Seleção de covariáveis usando o modelo de regressão de COX.....	34
Tabela 4.6 – Passo 3 - Seleção de covariáveis usando o modelo de regressão de COX.....	35
Tabela 4.7 – Passo 4 - Seleção de covariáveis usando o modelo de regressão de COX.....	35
Tabela 4.8 – Teste de proporcionalidade dos riscos no modelo de Cox	38
Tabela 4.9 – Estimativas da Função de Sobrevivência e da Função Taxa de Falha	39
Tabela 4.10 – Estimativas da Função de Sobrevivência (S0(t)) ajustadas para as covariáveis significativas	40
Tabela 4.11 – Estimativas da Função de Taxa de Falha ($\lambda_0(t)$) ajustadas para as covariáveis significativas	4Erro! Indicador não definido.

SUMÁRIO

1	INTRODUÇÃO	14
1.1	PREMISSAS E PROBLEMA DE PESQUISA	14
1.2	OBJETIVOS	16
1.3	JUSTIFICATIVA	16
1.4	PROCEDIMENTOS METODOLÓGICOS	17
1.5	ESTRUTURA DO TRABALHO	17
2	DEFESA DE CONSUMIDOR	18
2.1	DADOS DAS RECLAMAÇÕES	20
2.1.1	DADOS.....	20
2.1.2	DEFINIÇÃO DA VARIÁVEL RESPOSTA	21
2.1.3	PERÍODO DE AVALIAÇÃO	21
2.1.4	ANÁLISE DESCRITIVA DAS VARIÁVEIS.....	21
2.2	CONSIDERAÇÕES E SINTESE DO CAPÍTULO	24
3	REFERENCIAL TEÓRICO	25
3.1	ANÁLISE DE SOBREVIVÊNCIA	25
3.2	MODELO DE COX	27
4	MODELAGEM DAS RECLAMAÇÕES	28
4.1	ANÁLISE DOS DADOS	28
4.2	CONSIDERAÇÕES E TÉCNICAS DE TRATAMENTO DE DADOS APLICADAS	29
4.3	AJUSTE DO MODELO DE COX	32
4.4	RESULTADOS	36
4.5	INTERPRETAÇÃO DOS RESULTADOS DA APLICAÇÃO DO MODELO DE COX	39
4.6	DISCUSSÃO DOS RESULTADOS	43
5	CONSIDERAÇÕES FINAIS	45
5.1	SUGESTÕES PARA TRABALHOS FUTUROS.....	45

1 INTRODUÇÃO

O emprego de técnicas estatísticas para análise de dados cresceu substancialmente nas últimas décadas. Este crescimento é, em grande parte, atribuído à combinação do aprimoramento da própria técnica estatística e do desenvolvimento dos recursos computacionais. Dentre as técnicas estatísticas, a análise de sobrevivência está sendo largamente aplicada na medicina e na engenharia.

1.1 PREMISSAS E PROBLEMA DE PESQUISA

Segundo Bailar III e Mosteller (1992), os dois artigos mais citados na literatura estatística no período de 1987 e 1989 foram o do estimador de Kaplan-Meier para a função de sobrevivência (Kaplan e Meier, 1958) e o do modelo de Cox (Cox, 1972).

A análise de sobrevivência analisa e modela, geralmente, o tempo que leva para que eventos ocorram. A principal característica de dados de sobrevivência é a utilização da informação parcial da resposta, chamada de censura, a qual se refere a situações em que, por alguma razão, o acompanhamento do processo foi interrompido ou não foi preciso. Caso não existam censuras, pode-se utilizar as técnicas usuais de análise de variância (ANOVA) para fazer tal comparação. No entanto, na presença de censuras, tais técnicas não podem ser usadas, pois necessitam de todos os tempos de falha. Desta forma, faz-se necessário o uso dos métodos de análise de sobrevivência que integram na análise a informação contida nos dados censurados.

Primordialmente, o evento de falha citado é a morte, por isso o nome e a terminologia de “análise de sobrevivência”, contudo o âmbito da aplicação das técnicas utilizadas é muito mais amplo. Essencialmente, os mesmos métodos são empregados na análise de tempo de falha na engenharia, sendo chamada de análise de confiabilidade.

A partir dos anos 50 a confiabilidade dos produtos passou a ser estudada detalhadamente por meio dos testes de vida. Nos anos 70, novas técnicas de análise com várias aplicações foram introduzidas nas indústrias com o objetivo de

ajudar na solução dos problemas de confiabilidade. Em 1970, Nelson utilizou modelos estatísticos para trabalhar com a confiabilidade de produtos através do teste de vida acelerado, o qual envolve a aceleração de falhas submetendo o produto a condições estressantes com o objetivo de quantificar as suas características de vida relacionadas com condições de uso.

Embora existam métodos bem conhecidos para estimar distribuições de sobrevivência incondicionais, a modelagem de sobrevivência mais interessante examina a relação entre os tempos e as variáveis preditoras, geralmente denominadas covariáveis. O foco deste trabalho é o modelo de risco proporcionais introduzido por Cox (1972) em seu artigo no qual o estudo do tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse é ajustado por covariáveis. Por exemplo, o tempo de vida de um componente pode ser influenciado pela pressão exercida nele ou a temperatura a que está exposto ou, ainda, a comparação dos tempos de falha de dois grupos em que os pacientes são selecionados aleatoriamente para receber um determinado tratamento ou placebo.

Comumente a obtenção de uma única distribuição do tempo de sobrevivência não é suficiente para atender o objetivo da pesquisa e um aprofundamento se faz necessário com o intuito de comparar o tempo de sobrevivência de grupos distintos ou a verificação de outros fatores nesse tempo. Por exemplo, o tempo de vida de um componente pode ser influenciado pela pressão exercida nele ou a idade de um paciente pode influenciar no surgimento de determinada doença. Para satisfazer essa necessidade, incorpora-se o efeito dessas covariáveis no estudo utilizando um modelo estatístico de regressão. Há duas classes de modelos para isso, os modelos paramétricos e os semi-paramétricos. Essa segunda classe, introduzida por Cox (1972), é flexível e muito utilizada em dados de sobrevivência nos quais o efeito das covariáveis de interesse pode variar ao longo do tempo de duração do estudo.

Da aplicação dessas técnicas surge a questão direcionadora deste trabalho: Quais são os principais fatores que influenciam o atendimento a necessidades dos clientes nas reclamações registradas?

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Testar e avaliar se o modelo de Cox comumente utilizado para modelar tempo de vida é capaz de modelar com precisão ou de forma satisfatória o tempo de solução de conflitos utilizando uma ferramenta por meio da internet que permite a interlocução direta entre consumidores e empresas.

1.2.2 Objetivos Específicos

Este trabalho buscará atender os seguintes objetivos específicos para cumprimento do objetivo geral.

- Obter e ajustar os dados das reclamações;
- Demonstrar um método de análise descritiva de variáveis;
- Ajustar o modelo de Cox para as variáveis;
- Analisar os resíduos para observar a adequação do modelo ajustado;
- Apresentar os resultados finais;

1.3 JUSTIFICATIVA

Considerando a grande utilidade no uso de modelos de regressão na análise de dados censurados e, a real necessidade de analisar modelos contendo covariáveis dependentes no tempo pretende-se com este trabalho apresentar a aplicação do modelo de riscos proporcionais de Cox aos dados de um serviço público mantido pelo Ministério da Justiça e disponível em <https://www.consumidor.gov.br> para a solução alternativa de conflitos de consumo por meio da internet, que permite interlocução direta entre consumidores e empresas.

1.4 PROCEDIMENTOS METODOLÓGICOS

Este é um estudo comparativo de técnicas estatísticas de sobrevivência na presença de covariáveis que podem estar relacionadas com o tempo de resposta e apresenta a seguinte estratégia sequencial:

- Encontrar a distribuição de probabilidade que melhor modele os tempos de sobrevivência;
- Ajustar o modelo de Cox identificando as principais covariáveis (fatores) que afetam o tempo de sobrevivência;
- Discutir e Comparar os resultados, em especial a qualidade dos ajustes, acurácia das estimativas e diferentes encontradas;

1.5 ESTRUTURA DO TRABALHO

A presente monografia está dividida cinco capítulos. Uma descrição dos conceitos básicos necessários ao desenvolvimento deste trabalho é apresentada no Capítulo 2.

O Capítulo 3 descreve o modelo de riscos proporcionais de Cox, bem como os testes para avaliar o seu desempenho.

No Capítulo 4 são apresentados os detalhes dos dados e a aplicação do modelo descrito incluindo as covariáveis dependentes do tempo. Os dados referem-se ao acompanhamento de uma ferramenta, concebida com base em princípios de transparência e controle social, a qual através de uma plataforma de informação resolução de conflito fornecendo ao Estado informações essenciais à elaboração e implementação de políticas públicas de defesa dos consumidores e incentivando a competitividade no mercado pela melhoria da qualidade e do atendimento ao consumidor. As conclusões estatísticas são apresentadas no Capítulo 5.

Para a realização deste trabalho foram utilizados os softwares R. O software R é interpretado como uma linguagem computacional designada para análise de dados estatísticos com grande flexibilidade, incluindo uma ampla variedade de métodos estatísticos. Uma característica importante é que o R é um software gratuito.

2 DEFESA DO CONSUMIDOR

Este capítulo apresenta um breve histórico da defesa do consumidor do Brasil, seu embasamento jurídico e constitucional e uma ferramenta virtual que visa facilitar o entendimento entre fornecedores e consumidores em casos de conflito. Apresenta a origem dos dados estudados, bem como os parâmetros para a geração destes.

O início da defesa do consumidor no Brasil deu-se nos anos 60, a partir do reconhecimento, pelo então presidente americano, da vulnerabilidade do consumidor e sua importância nas relações comerciais nos Estados Unidos, em uma mensagem ao Congresso Nacional estadunidense, quando esse afirmou o caráter universal da proteção dos direitos dos consumidores, tais como o direito à segurança, a informação e de escolha.

Com o advento da Constituição da República Federativa do Brasil em 1988, a Constituição Cidadã, consagrou-se a proteção do consumidor como direito fundamental e princípio da ordem econômica (art. 5º, XXXII, e 170, V), cabendo ao Estado a promoção da defesa do consumidor, na forma da lei.

A luz da constituição, foi elaborado o Código de Defesa do Consumidor, por meio da Lei nº 8.078, de 1990, a fim de estabelecer normas de proteção e defesa do consumidor e mitigar as fraquezas e deficiências que haviam nas relações de consumo, assegurando o equilíbrio de direitos e obrigações de parte a parte.

Em 2012, por meio do Decreto nº 7.738, foi criada a Secretaria Nacional do Consumidor, à qual cabe exercer as competências estabelecidas no Código de Defesa do Consumidor tais como formular, promover, supervisionar e coordenar a Política Nacional de Proteção e Defesa do Consumidor.

O consumidor.gov.br é um serviço público para solução alternativa de conflitos de consumo, disponibilizado por meio de plataforma tecnológica de informação, interação e compartilhamento de dados monitorada pelos órgãos de defesa do consumidor e pela Secretaria Nacional do Consumidor do Ministério da Justiça, com o apoio da sociedade.

Lançado em 2014, o Consumidor gov.br é um espaço público de atendimento ao consumidor, promovendo o diálogo efetivo entre consumidores e empresas de diversos segmentos.

Trata-se de uma plataforma tecnológica de informação, interação e compartilhamento de dados, monitorada pelos PROCONs e pela Secretaria Nacional do Consumidor do Ministério da Justiça, com o apoio da sociedade.

A ferramenta, concebida com base em princípios de transparência e controle social, fornece ao Estado informações essenciais à elaboração e implementação de políticas públicas de defesa dos consumidores e incentiva a competitividade no mercado pela melhoria da qualidade e do atendimento ao consumidor.

O Consumidor.gov.br coloca as relações entre Consumidores, Fornecedores e o Estado em um novo patamar, a partir das seguintes premissas:

- I. Transparência e controle social são imprescindíveis à efetividade dos direitos;
- II. As informações apresentadas pelos cidadãos consumidores são estratégicas para gestão e execução de políticas públicas de defesa do consumidor; e
- III. O acesso a informação potencializa o poder de escolha dos consumidores e contribui para o aprimoramento das relações de consumo.

A participação das empresas na plataforma é voluntária e se comprometem em conhecer, analisar e investir todos os esforços disponíveis para a solução dos problemas apresentados e responder as reclamações de seus consumidores em até 10 dias. O consumidor, por sua vez, deve se identificar adequadamente e comprometer-se a apresentar todos os dados e informações relativas à reclamação relatada.

O desempenho das empresas participantes pode ser monitorado por meio de indicadores a respeito da quantidade de reclamações finalizadas por empresa, o índice de resolução, a satisfação do consumidor com o atendimento dado e o prazo médio de resposta.

A Secretaria Nacional do Consumidor do Ministério da Justiça é a responsável pela gestão, disponibilização e manutenção da ferramenta, bem como pela articulação com demais órgãos e entidades do Sistema Nacional de Defesa do Consumidor que por meio de cooperação técnica, apoiam e atuam na consecução dos objetivos do serviço.

A criação dessa plataforma guardou relação com o disposto no artigo 4º inciso V, da Lei 8.078/1990 e artigo 7º, incisos I, II e III, do Decreto 7.963/2013.

2.1 DADOS DAS RECLAMAÇÕES

Os dados utilizados para o desenvolvimento desse estudo contemplam as informações das reclamações registradas na ferramenta Consumidor.gov.br, tendo sido obtidos no portal brasileiro de dados abertos (www.dados.gov.br).

2.1.1 Dados

A seguinte tabela mostra um resumo dos principais dados utilizados na análise.

Tabela 2.1 - Variáveis coletadas das reclamações

Dado	Descrição	Variável
Região	Sigla da região geográfica do consumidor reclamante	RG
Sexo	Sigla do sexo do consumidor reclamante	SX
Faixa Etária	Faixa etária do consumidor	FE
Ano Abertura	Ano de abertura da reclamação pelo consumidor	AA
Mês Abertura	Número do mês de abertura da reclamação pelo consumidor	MA
Data Abertura	Data de abertura da reclamação pelo consumidor	DA
Data Resposta	Data de resposta da reclamação pela empresa	DR
Data Finalização	Data de finalização da reclamação	DF
Tempo Resposta	Número de dias para a resposta da reclamação pela empresa.	TR
Respondida	Sigla que indica se a empresa respondeu reclamação ou não	RE
Situação	Situação atual da reclamação no sistema	SI
Avaliação Reclamação	Classificação atribuída pelo consumidor sobre o desfecho da reclamações	ARec

Fonte: Portal Brasileiro de Dados Abertos (www.dados.gov.br)

As variáveis apresentadas foram analisadas a fim de identificar a suas relações com a variável resposta. Para fins desse estudo primário foram selecionadas somente variáveis de perfil demográfico, considerando que não há informação suficiente sobre quais seriam as principais características das reclamações que poderiam afetar a resolução das mesmas.

2.1.2 Definição da Variável Resposta

A definição da Variável Resposta (falha) representa um passo extremamente importante, pois é ela que será modelada. O presente trabalho estudará a probabilidade da solução das reclamações registradas por meio dessa ferramenta dentro do tempo acordado de 10 dias e identificará se há alguma dependência entre essa solução e características da reclamação. Para isso o evento "falha" foi definido como as "reclamações respondidas e não resolvidas" ou das "reclamações não respondidas" e "não avaliadas ou não resolvidas".

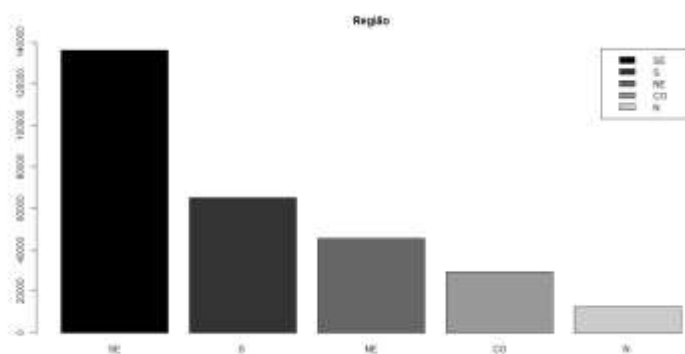
2.1.3 Período de Avaliação

A construção da base de dados utiliza informações das reclamações registradas na ferramenta em 2016 e disponibilizadas no Portal Brasileiro de Dados Abertos.

2.1.4 Análise Descritiva das Variáveis

A etapa inicial de análise dos dados chamada de análise descritiva ou univariada ou bivariada é utilizada para descrever e resumir os dados, assim com identificar possíveis faltas ou deficiências nestes. Nas figuras abaixo observa-se como os dados se comportam com relação à concentração de classe, dados faltantes, média, mínimo, máximo entre outras, para o período de Jan/16 – Dez/16.

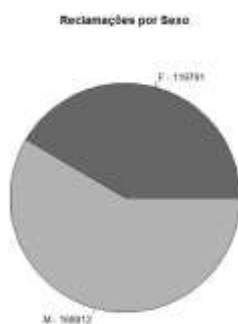
Figura 2.1 - Distribuição das reclamações por Região



Fonte: o autor (2017).

Nas Figuras 2.1 note que há uma concentração nos estados das regiões sudeste e sul as quais concentram aproximadamente 69% das reclamações.

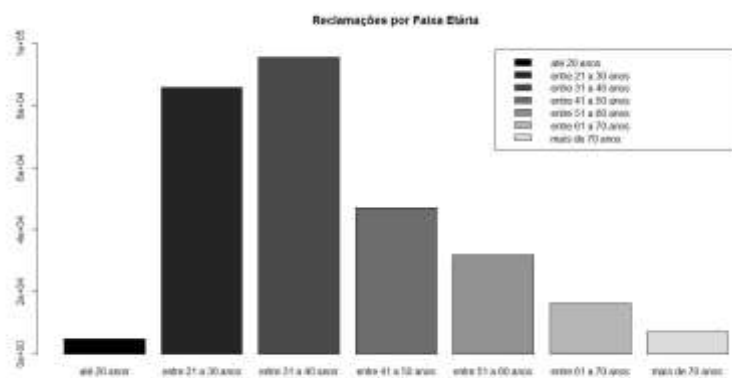
Figura 2.2 - Reclamações por Sexo



Fonte: o autor (2017).

Na Figura 2.2 note que aproximadamente 59% das reclamações foram feitas por homens.

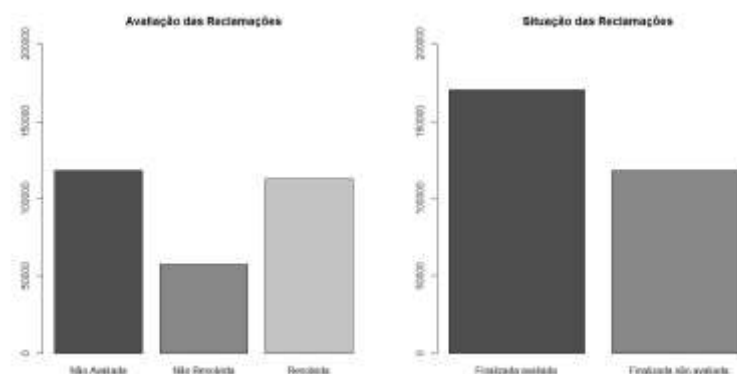
Figura 2.3 - Faixa Etária dos Reclamantes



Fonte: o autor (2017).

Na Figura 2.3 observe que curiosamente as reclamações foram realizadas por pessoas majoritariamente com mais de 21 e menos de 40 anos.

Figura 2.4 - Situação e Avaliação das Reclamações



Fonte: o autor (2017).

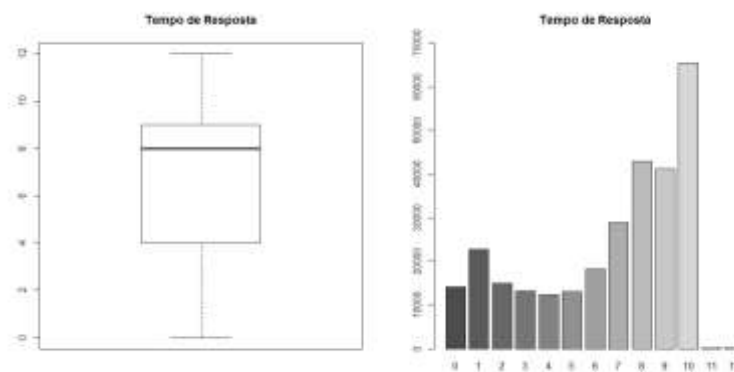
Figura 2.5 - Situação e Avaliação das Reclamações Cruzadas



Fonte: o autor (2017).

Nas Figuras 2.4 e 2.5 observe que existe uma equivalência entre as variáveis Situação e Avaliação das Reclamações no que tange a classe Finalizada e Não Avaliada. Isso é de particular interesse para a análise, pois para aproximadamente 41% das reclamações estão foram Finalizadas e Não avaliadas. Quando uma reclamação não é avaliada fica prejudicada a identificação da nossa variável de interesse.

Figura 2.6 – Tempo de Resposta



Fonte: o autor (2017).

Na Figura 2.6 note que o tempo mediano de resposta são 8 dias variando entre 0 e 12 dias ultrapassando, em alguns casos, o objetivo do programa que é a resposta em até 10 dias.

2.2 CONSIDERAÇÕES E SINTESE DO CAPITULO

Neste capítulo foram abordados os principais conceitos relacionados à defesa do consumidor e de interesse ao objetivo deste trabalho, citados no capítulo 1. Também foi apresentada a origem do direito do consumidor e seu embasamento jurídico, bem como a ferramenta de solução alternativa de conflitos de consumo. Adicionalmente, foram apresentados dados por meio de uma análise exploratória das variáveis, assim como a definição do evento de interesse (falha).

No próximo capítulo serão abordados os aspectos teóricos das técnicas utilizadas para a identificação das principais covariáveis e sua relação com a falha.

3 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os conceitos técnicos inerentes ao processo de análise dos dados aplicados neste trabalho, tendo como foco nas técnicas estatísticas a serem aplicadas e uma breve fundamentação teórica para a aplicação das mesmas na análise. Os itens abordados apresentam a justificativa do uso das referidas técnicas e da escolha pelo modelo de COX.

As técnicas estatísticas tradicionais de regressão linear não são aplicáveis para modelar o tempo de sobrevivência como uma função de um conjunto de variáveis preditoras, pois os tempos de sobrevivência são tipicamente números positivos e a regressão linear ordinária pode não ser a melhor escolha, a menos que esses tempos sejam transformados de uma maneira que elimine essa restrição. Adicionalmente, e mais importante, a regressão linear comum não pode efetivamente lidar com a censura de observações.

Na regressão logística, o interesse é estudar como fatores de risco estão associados à presença ou ausência de doença. Às vezes, no entanto, o interesse é como um fator de risco ou tratamento afeta o tempo para a doença ou algum outro evento. Ou podemos ter abandono do estudo e, portanto, sujeitos que não temos certeza de ter doença ou não. Nesses casos, a regressão logística não é apropriada.

3.1 ANÁLISE DE SOBREVIVÊNCIA

A análise de sobrevivência é, em geral, utilizada quando há interesse no tempo até a ocorrência de um evento de interesse. Este evento pode ser a morte, o casamento, a falência de uma empresa ou a falha de um sistema. O tempo até esse evento pode ser medido em dias, semanas, ciclos, quilômetros ou etc.

As observações são chamadas censuradas quando a informação sobre o tempo de sobrevivência está incompleta; A forma mais comum é a censura a direita onde, por exemplo, um paciente não experimenta o evento de interesse durante a duração de um estudo de 20 semanas. Outro exemplo de censura a direita é quando um paciente abandona o estudo antes do final do tempo de observação e não experimentou o evento. A censura é um dos temas na análise de sobrevivência, representando um tipo particular de dados faltantes (missing data).

Ao contrário dos modelos de regressão ordinários, os métodos de sobrevivência incorporam corretamente informações de observações censuradas e não censuradas na estimativa de parâmetros do modelo. A variável dependente na análise de sobrevivência é composta por duas partes: uma é o momento do evento e a outra é o status do evento, que registra se o evento de interesse ocorreu ou não. Pode-se então estimar duas funções que dependem do tempo e do risco. As funções de sobrevivência e risco são conceitos-chave na análise de sobrevivência para descrever a distribuição de eventos. A função de sobrevivência dá, para cada momento, a probabilidade de sobreviver (ou não) até determinado momento. A função de *hazard* dá o potencial de que o evento ocorra, por unidade de tempo, dado que um indivíduo tenha sobrevivido até o tempo especificado. Embora esses sejam frequentemente de interesse indireto, muitas outras quantidades de interesse (e.g. sobrevivência mediana) podem ser estimadas posteriormente sabendo a função de risco ou, propriamente, de sobrevivência.

É geralmente de interesse em estudos de sobrevivência descrever o relacionamento de um fator de interesse (por exemplo, tratamento) até o momento do evento, na presença de várias covariáveis, como idade, gênero, raça, etc. Uma série de modelos estão disponíveis para analisar a relação de um conjunto de variáveis preditoras com o tempo de sobrevivência. Os métodos incluem abordagens paramétricas, não paramétricas e semi-paramétricas.

Os métodos paramétricos assumem que a distribuição subjacente dos tempos de sobrevivência segue certas distribuições de probabilidade conhecidas. Os mais populares incluem as distribuições Exponencial, Weibull e Lognormal. A descrição da distribuição dos tempos de sobrevivência e a mudança na sua distribuição em função dos dados são de interesse. Os parâmetros do modelo nessas configurações geralmente são estimados usando uma modificação apropriada da máxima verossimilhança.

Um estimador não paramétrico da função de sobrevivência, o método de Kaplan Meier é amplamente utilizado para estimar e traçar probabilidades de sobrevivência em função do tempo. Ele pode ser usado para obter estatísticas descritivas univariadas para dados de sobrevivência, incluindo o tempo médio de sobrevivência, e comparar a experiência de sobrevivência para dois ou mais grupos de indivíduos. Para testar as diferenças globais entre as curvas de sobrevivência estimadas de dois ou mais grupos de indivíduos, como os machos versus fêmeas,

ou os grupos tratados versus não tratados (controle), vários testes estão disponíveis incluindo o teste log-rank. Isso pode ser motivado como um tipo de teste qui-quadrado um teste amplamente utilizado na prática, e na realidade é um método para comparar as curvas de Kaplan-Meier estimadas para cada grupo de indivíduos.

Um modelo de regressão popular para a análise de dados de sobrevivência é o modelo de regressão de riscos proporcionais de Cox, o qual permite testar as diferenças nos tempos de sobrevivência de dois ou mais grupos de interesse permitindo ajustar as covariáveis de interesse. O modelo de regressão Cox é um modelo semi-paramétrico, fazendo menos suposições do que os métodos paramétricos típicos, mas com mais pressupostos do que os métodos não paramétricos descritos acima. Em particular, e em contraste com modelos paramétricos, não faz suposições sobre a forma da chamada função de risco de base.

3.2 MODELO DE COX

O Modelo de Riscos Proporcionais de Cox utilizado na Análise de Sobrevivência (Survival Analysis) possui larga aplicação na área biomédica e se diferencia daqueles estimados a partir de técnicas logit, probit e análise discriminante, uma vez que fornece não apenas a probabilidade de que um determinado evento ocorra no futuro, mas também uma estimativa do tempo até sua ocorrência.

Nesse sentido, a estimação da probabilidade de sobrevivência para diferentes horizontes de tempo permite obter a efetividade da resolução dos conflitos entre as empresas e os consumidores usuários da ferramenta.

Entretanto, o modelo possui algumas limitações. Da mesma forma que outras técnicas estatísticas utilizadas na previsão de eventos, o modelo de risco proporcional exige dados anteriores à falha sob análise. Além disso, o modelo assume valores das variáveis independentes não se alteram ao longo do que o período de tempo no qual se desenvolve o estudo (suposição de proporcionalidade das taxas de falha). Segundo Whalen (1991; p.23), a violação dessa hipótese pode reduzir a precisão do modelo.

4 MODELAGEM DAS RECLAMAÇÕES

Este capítulo aborda a aplicação do modelo de Regressão de Cox, o qual permite a análise dos dados em questão, pois nosso interesse é determinar a confiabilidade da ação do site sobre a solução das reclamações dos clientes medida por meio da resposta satisfatória das empresas sob o ponto de vista dos clientes, ajustando por covariáveis que possam impactar na resolução ou não dessas reclamações. Adicionalmente, serão abordados os dados no que tange a análise e as técnicas aplicadas nestes.

4.1 ANÁLISE DOS DADOS

Conforme item 2.1.2, a variável resposta (falha) foi definida como as "reclamações respondidas e não resolvidas" ou das "reclamações não respondidas" e "não avaliadas ou não resolvidas". Entretanto, as figuras 2.8 e 2.9 observou-se a classe Finalizada e Não Avaliada impede a determinação da variável resposta. Além disso, considerar a reclamações não respondidas na definição da variável resposta é temerário, pois a não é possível saber se realmente as reclamações foram ou não resolvidas.

Tabela 4.1 - Número de Reclamações com Avaliação Recebida vs Resposta da Empresa

Avaliação Recebida	Resposta Empresa		Total Respondido
	Sim	Não	
Resolvida	112.731	163	112.894
Não Resolvida	57.238	324	57.562
Não Avaliada	117.462	685	118.147
Total Avaliado	287.431	1.172	288.603

Fonte: o autor (2017).

Observe que na tabela acima temos 117.462 reclamações não avaliadas e 1.172 que não foram respondidas. Considerando o número restante dos dados e o fato que a não divulgação desses dados impacta diretamente na construção da variável resposta, decidiu-se excluir esses dados da análise.

Tabela 4.2 - Número de Reclamações com Avaliação Recebida vs Resposta da Empresa já considerando as exclusões supramencionadas.

Avaliação Recebida	Resposta Empresa		
	Sim	Não	Total Respondido
Resolvida	112.731	-	11.731
Não Resolvida	57.238	-	57.238
Não Avaliada	-	-	-
Total Avaliado	169.969	-	169.969

Note que na tabela 4.2 restaram 169.969 reclamações a serem avaliadas.

4.2 CONSIDERAÇÕES E TÉCNICAS DE TRATAMENTO DE DADOS APLICADAS

Considerando as três covariáveis (Região, Sexo e Faixa Etária) potencialmente importantes para descrever o comportamento da resposta, existem diversos modelos que poderiam ser resultantes das combinações das classes de cada variável. Sendo assim foram utilizadas as curvas de Kaplan-Meier para comparar as covariáveis e as classes dentro de cada uma delas.

4.3.1 Variáveis Dummies

Nesse sentido foram criadas variáveis auxiliares que representam as classes dentro de cada uma das variáveis. Por meio de variáveis *dummy* é possível usar variáveis categóricas para a estimação de modelos. Elas podem ser utilizadas em modelos simples, em que a variável dá origem a apenas uma *dummy*; ou complexos, no qual a variável qualitativa origina um vetor com duas ou mais *dummies* (Missio & Jacobi, 2007).

O quadro 4.1 relaciona as variáveis auxiliares/artificiais (*dummies*) criadas e a covariável / classe que ela representa.

Quadro 4.1 – Variáveis Dummies para cada classe

Covariável	Classe	Variável Dummy
Região (RG)	CO	RG_V1
	N	RG_V2
	NE	RG_V3
	S	RG_V4
	SE	-
Sexo (SX)	M	SX_V1
	F	-
Faixa Etária (FE)	até 20 anos	FE_V1
	entre 21 a 30 anos	-
	entre 31 a 40 anos	FE_V2
	entre 41 a 50 anos	FE_V3
	entre 51 a 60 anos	FE_V4
	entre 61 a 70 anos	FE_V5
mais de 70 anos	FE_V6	

Fonte: Autor (2017).

Variáveis dummy indicam presença ou ausência de atributos transformando as variáveis originais em artificiais, as quais assumem valores de 0 ou 1, onde zero indica ausência de atributo e um, sua presença. Para codificações com mais de duas categorias é recomendado utilizar sempre uma variável a menos; isto é, se a variável qualitativa pode assumir N valores, deverá ser representada por um vetor de N-1 dummies.

Tabela 4.3 - RG_V1 vs Covariável Região.

RG	RG_V1	
	0	1
CO	0	17287
N	6730	0
NE	27706	0
S	35227	0
SE	83019	0

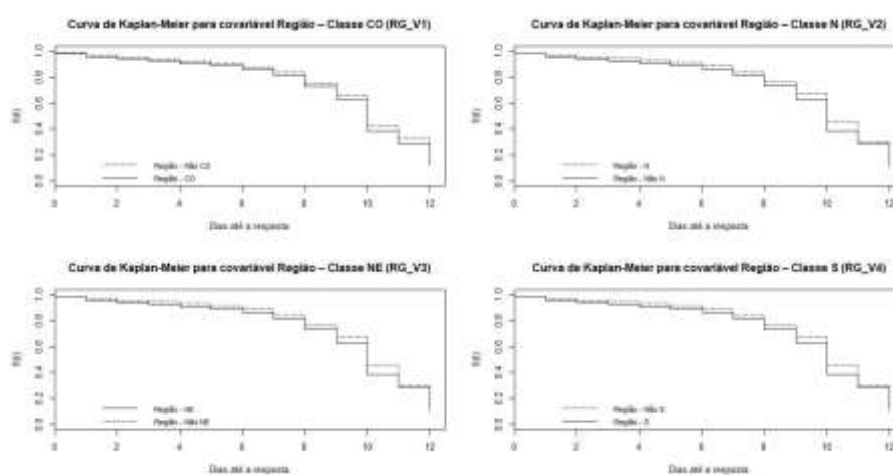
Fonte: Autor (2017).

Observe que a variável *dummy* RG_V1 na tabela 4.3 indica a existência (RG_V1 = 1) do atributo para 17287 registros e a não existência (RG_V1 = 0) para os demais.

4.3.2 Análise Descritiva e Exploratória

Nesta etapa serão utilizados métodos não paramétricos para a análise das onze variáveis dummies que representam as classes das covariáveis em estudo. Como as variáveis são dicotômicas é possível construir as estimativas de Kaplan-Meier para essas categorias. As figuras a seguir conterão as curvas de Kaplan-Meier para cada variável auxiliar.

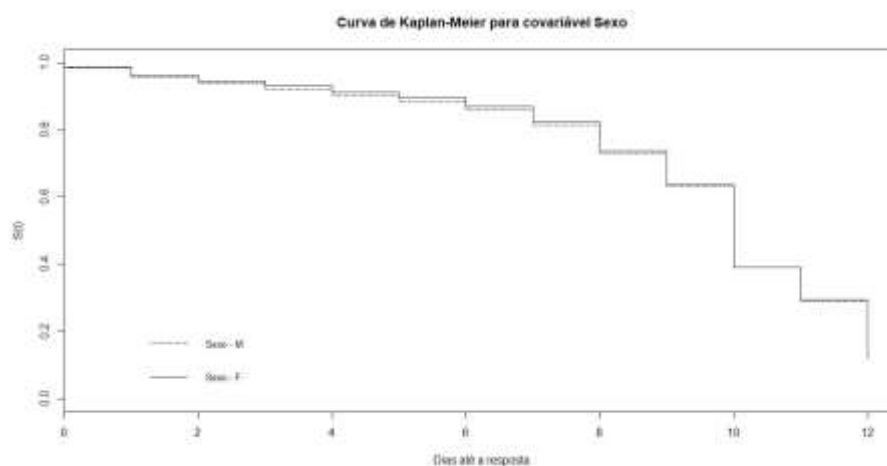
Figura 4.1 – Curva de Kaplan-Meier para covariável Região



Fonte: o autor (2017).

Na Figura 4.1 note que as regiões se diferenciam consistentemente conforme a curva de Kaplan-Meier.

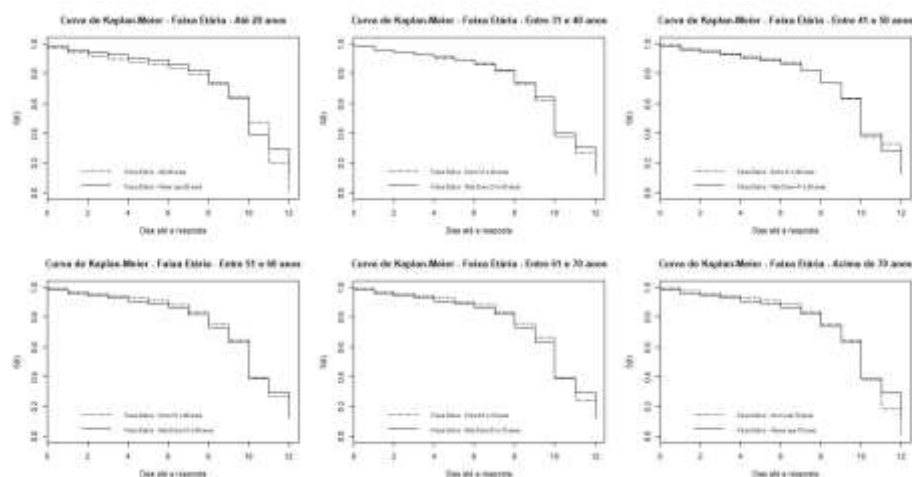
Figura 4.2 – Curva de Kaplan-Meier para covariável Sexo



Fonte: o autor (2017).

Na Figura 4.2 observe que os homens diferem ligeiramente das mulheres em termos das reclamações.

Figura 4.3 – Curva de Kaplan-Meier para covariável Faixa Etária



Fonte: o autor (2017).

Na Figura 4.3 note que as classes da covariável faixa etária se cruzam diversas vezes. Este fato sugere que a influência da variável na resposta não é significativa e, também, pode não estar sendo atendida a suposição de riscos proporcionais para que o modelo de Cox seja considerado adequado para estes dados. A significância e o atendimento dessas suposições serão verificados nos capítulos subsequente durante as fases de ajuste do modelo de Cox.

4.3 AJUSTE DO MODELO DE COX

O modelo de regressão de Cox é caracterizado pelos coeficientes betas, que medem os efeitos das covariáveis sobre a função de taxa de falha. Estas estimativas são obtidas a partir das observações amostrais para que o modelo fique determinado. Neste sentido, o método de estimação utilizado foi o de máxima verossimilhança parcial para se fazer inferências acerca dos parâmetros do modelo. Este método consiste em condicionar a construção da função de verossimilhança aos dados de falha e de censura.

4.3.1 Estratégia para a Seleção de Variáveis

Considerando que as onze variáveis dummies, as quais representam as n-1 classes de cada covariável, são potencialmente relevantes para descrever o comportamento da resposta e, portanto, um número impraticável de possíveis modelos pode ser ajustado a fim de ser selecionado o melhor ou o com menor desvio, deve-se aplicar uma estratégia para a seleção de variáveis. A estratégia optada é derivada da proposta de Collett (1994) e consistem nos passos subsequentes.

4.3.1.1 Passo 1

Ajustar todos os modelos contendo uma única covariável (i.e. variável, dummy no caso prático em questão). Incluir todas as covariáveis que forem significativas ao nível de 0.10.

Tabela 4.4 - Passo 1 - Seleção de covariáveis usando o modelo de regressão de COX.

Variáveis Ajustadas	-2 log L(Theta)	TRV	Valor p
Nulo	1290289	-	-
RG_V1	1290230	59	0,0000
RG_V2	1290246	43	0,0000
RG_V3	1290278	11	0,0009
RG_V4	1290285	4	0,0455
SX_V1	1290283	6	0,0143
FE_V1	1290289	0	1,0000
FE_V2	1290257	32	0,0000
FE_V3	1290288	1	0,3173
FE_V4	1290270	19	0,0000
FE_V5	1290276	13	0,0003
FE_V6	1290287	2	0,1573

Fonte: Autor (2017).

Observe que as variáveis FE_V1, FE_V3 e FE_V6 que representam as classes “Até 20 anos”, “Entre 41 e 50 anos” e “Mais de 70 anos” da covariável Faixa

Etária respectivamente, não são significativas ao nível de 0.10 e, possivelmente, serão excluídas.

4.3.1.2 Passo 2

Ajustar conjuntamente as covariáveis significativas do passo 1. Ajustar modelos reduzidos retirando uma única variável de cada vez. Somente aquelas que atingirem a significância permaneceram no modelo.

Tabela 4.5 - Passo 2 - Seleção de covariáveis usando o modelo de regressão de COX.

Variáveis Ajustadas	-2 log L(Theta)	TRV	Valor p
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290102	-	-
RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290179	77	0,0000
RG_V1+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290158	56	0,0000
RG_V1+RG_V2+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290102	0	1,0000
RG_V1+RG_V2+RG_V3+SX_V1+FE_V2+FE_V4+FE_V5	1290117	15	0,0001
RG_V1+RG_V2+RG_V3+RG_V4+FE_V2+FE_V4+FE_V5	1290106	4	0,0455
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V4+FE_V5	1290118	16	0,0001
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V5	1290113	11	0,0009
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4	1290113	11	0,0009

Fonte: Autor (2017).

A variável RG_V3, a qual representa a região nordeste (NE), não é significativa ao nível de 0.10 e, possivelmente, será excluída.

4.3.1.3 Passo 3

Ajustar novos modelos incluindo as variáveis retidas no passo 2 uma a uma a fim de confirmar a não significância das mesmas. Entretanto somente a variável RG_V3 foi excluída no passo 2 então não se faz necessário o ajuste de novos modelos, pois os dados já foram obtidos.

Tabela 4.6 - Passo 3 - Seleção de covariáveis usando o modelo de regressão de COX.

Variáveis Ajustadas	-2 log L(Theta)	TRV	Valor p
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290102	-	-
RG_V1+RG_V2+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290102	0	1,0000

Fonte: Autor (2017).

A tabela 4.6 apenas expõe claramente que a variável RG_V3, a qual representa a região nordeste (NE), não é significativa ao nível de 0.10 e, possivelmente, será excluída.

4.3.1.4 Passo 4

Ajustar as variáveis significativas do passo 3 juntamente com as excluídas no passo 2 e adicionar uma a uma as variáveis excluídas no passo 1 para confirmar a não significância estatística das mesmas.

Tabela 4.7 - Passo 4 - Seleção de covariáveis usando o modelo de regressão de COX.

Variáveis Ajustadas	-2 log L(Theta)	TRV	Valor p	GL
RG_V1+RG_V2+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290102	-	-	-
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5	1290102	0	1,0000	1
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5+FE_V1	1290101	1	0,6065	2
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5+FE_V3	1290098	4	0,1353	2
RG_V1+RG_V2+RG_V3+RG_V4+SX_V1+FE_V2+FE_V4+FE_V5+FE_V6	1290100	2	0,3679	2

Fonte: Autor (2017).

As variáveis provaram-se não serem significativas e serão excluídas. Neste passo é importante lembrar que o teste Qui-Quadrado utilizado para testar a razão de verossimilhanças (TRV) deve utilizar como graus de liberdade a diferença entre o número de parâmetros dos TRVs.

4.3.1.5 Passo 5

Ajustar novos modelos incluindo alguma variável significativa identificada no passo 4 e, posteriormente, modelos excluindo uma a uma as variáveis para observar possibilidades de redução do modelo.

Neste caso prático, não foi identificada nenhuma variável significativa nos passos 3 e 4, então os modelos ajustados aqui são os mesmos do passo 2 onde foram verificadas as significâncias de cada uma das variáveis. Para maiores detalhes ver Tabela 4.5.

4.3.1.6 Passo 6

Ajustar o modelo final utilizando as variáveis sobreviventes do passo 5.

Quadro 4.2 – Modelo Final Antes da avaliação da suposição de riscos proporcionais

Variável	Coeficiente	Exp(Coef)	Int Inf 0.95	Int Sup 0.95
RG_V1	-0,1301	0,8780	0,8531	0,9036
RG_V2	-0,1727	0,8414	0,8036	0,8810
RG_V4	-0,0421	0,9588	0,9393	0,9787
SX_V1	0,0181	1,0182	1,0012	1,0356
FE_V2	0,0375	1,0383	1,0197	1,0571
FE_V4	-0,0476	0,9535	0,9272	0,9805
FE_V5	-0,0650	0,9371	0,9012	0,9744

Fonte: Autor (2017).

Depois do processo de seleção de variáveis, restaram aquelas que são realmente importantes para explicar a resposta. O modelo final, antes da avaliação da suposição de riscos proporcionais, com os efeitos principais e as interações significativas estão apresentados no quadro 4.2.

4.4 RESULTADOS

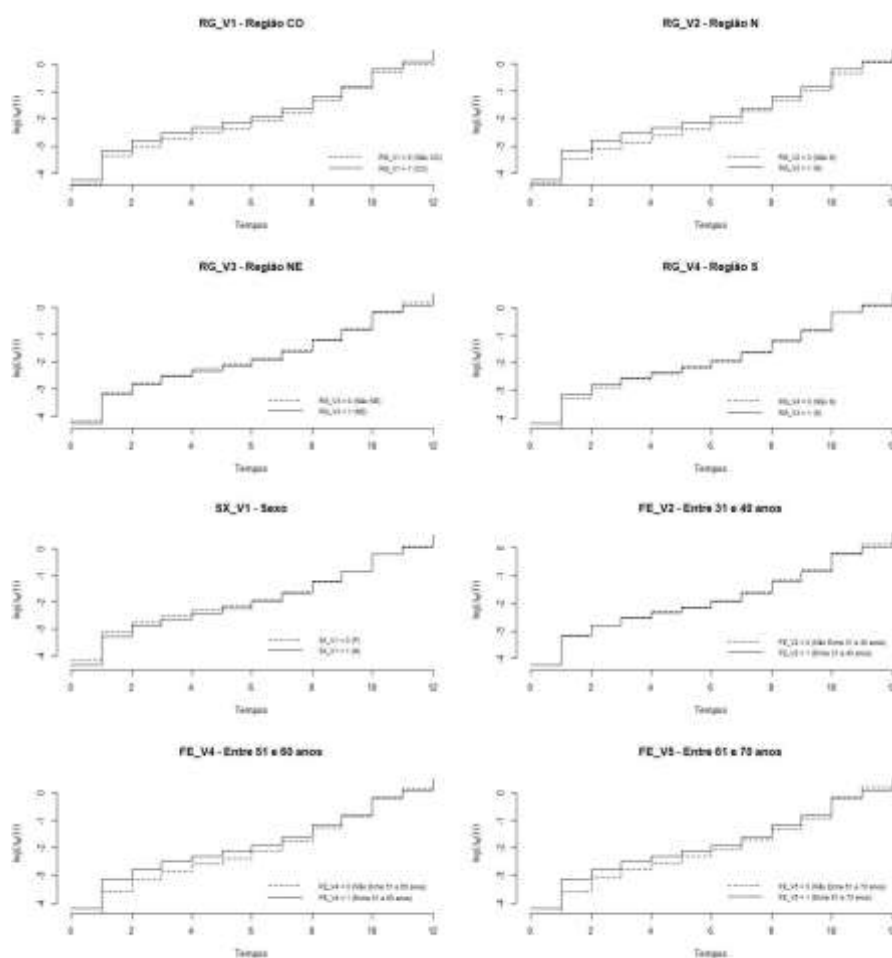
Não obstante o modelo tenha sido estimado, ainda é necessária a verificação o atendimento da suposição dos riscos proporcionais para que o modelo de Cox seja considerado adequado. De acordo com Colosimo e Giolo (2006),

técnicas gráficas aplicadas aos diferentes resíduos obtidos na regressão são particularmente muito utilizadas para examinar a distribuição dos erros. Estas técnicas, como observados em Klein e Moeschberger (1997), devem ser utilizadas primariamente para rejeitar modelos inadequados e não para “provar” que um determinado modelo é correto.

Os métodos gráficos aplicados ao modelo em questão utilizaram o logaritmo da função de risco acumulada e os resíduos padronizados de Schoenfeld. Ambos devem ser construídos para cada variável do modelo.

4.4.1 Logaritmo da função de risco acumulada versus tempo para as covariáveis

Figura 4.4 – Suposição de riscos proporcionais para as covariáveis utilizando gráficos do logaritmo da função de risco acumulada versus tempo para as covariáveis

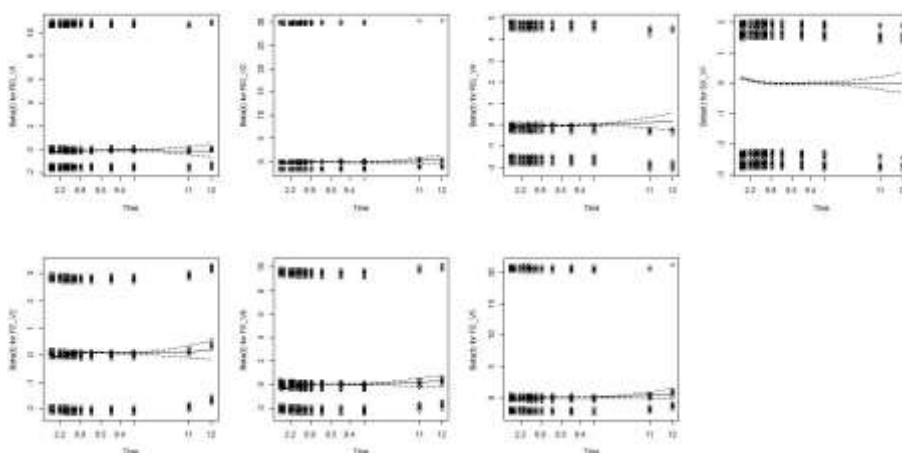


Aparentemente não há evidência de tendência ao longo do tempo no logaritmo da função de risco apresentados na análise gráfica, pois embora não

totalmente regular as linhas não se cruzam o que infere um aparente atendimento à suposição.

4.4.2 Suposição de riscos proporcionais para as covariáveis utilizando os resíduos de Schoenfeld

Figura 4.5 – Gráfico dos resíduos de Schoenfeld para cada variável



Fonte: Autor (2017).

Aparentemente não há evidência de tendência ao longo do tempo nos resíduos apresentados na análise gráfica implicando em um aparente atendimento à suposição. Contudo o teste de proporcionalidade dos riscos contido na Tabela 4.9 indica que há evidência da quebra dessa suposição, porém considerando que a amostra é grande essa quebra pode ser irreal decorrente do tamanho amostral.

Tabela 4.8 - Teste de proporcionalidade dos riscos no modelo de Cox.

Variável	rho	chisq	p
RG_V1	0,0185	19,5900	0,0000
RG_V2	0,0118	7,9900	0,0047
RG_V4	0,0166	15,7900	0,0001
SX_V1	-0,0246	34,6000	0,0000
FE_V2	0,0231	30,5800	0,0000
FE_V4	0,0418	99,9000	0,0000
FE_V5	0,0287	46,9300	0,0000

Fonte: Autor (2017).

O teste de proporcionalidade dos riscos contido na Tabela 4.8 indica que há evidência da quebra dessa suposição dado o valor p baixo, porém considerando que a amostra é grande essa quebra pode ser irreal decorrente do tamanho amostral considerando o rho próximo de zero.

4.5 INTERPRETAÇÃO DOS RESULTADOS DA APLICAÇÃO DO MODELO DE COX

Para se obter as estimativas da função de sobrevivência considerando cada covariável é necessário primariamente o cálculo da função de sobrevivência para toda a amostra.

Tabela 4.9 – Estimativas da Função de Sobrevivência ($S_0(t)$) e da Função Taxa de Falha ($\lambda_0(t)$)

Tempo	$S_0(t)$	$\lambda_0(t)$
0	0,9855	0,0146
1	0,9587	0,0422
2	0,9410	0,0608
3	0,9238	0,0792
4	0,9068	0,0978
5	0,8897	0,1169
6	0,8641	0,1461
7	0,8190	0,1997
8	0,7372	0,3049
9	0,6425	0,4424
10	0,4357	0,8309
11	0,3369	1,0879
12	0,1908	1,6565

Fonte: Autor (2017).

Observe que na Tabela 4.9 são apresentadas as estimativas da função de sobrevivência e da taxa de falhas. Neste sentido, pode-se afirmar que com 10 dias 43% das reclamações serão respondidas e resolvidas. É notável que a taxa de falhas aumenta com o passar do tempo, este fato pode ser originário de diferentes problemas que implicam na não resolução da reclamação.

Não obstante o foco da utilização do modelo de Cox neste estudo deve-se ao interesse de saber se há alguma característica demográfica que está ligada a resolução das reclamações. Neste sentido, deve-se ajustar as funções de Sobrevivência e de Risco estimadas sensibilizando-as para cada covariável significativa. Em termos gerais, pode-se expressar as funções supramencionadas por:

- Função de Sobrevivência

$$S(t | x) = [S_0(t)]^{\exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n\}};$$

e

- Função de Taxa de Falha

$$\lambda(t | x) = \lambda_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n\};$$

em que n é o número de parâmetros estimados pelo modelo de Cox.

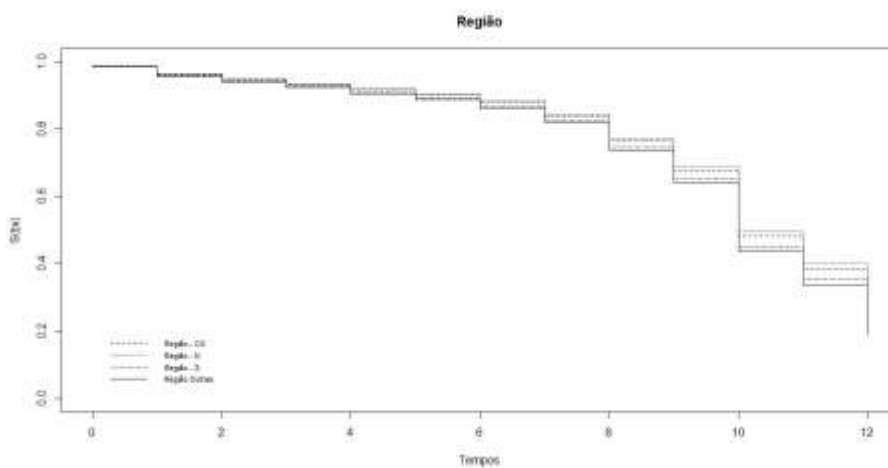
Tabela 4.10 – Estimativas da Função de Sobrevivência ($S_0(t)$) ajustadas para as covariáveis significativas

Tempo	S0(t)	RG_V1	RG_V2	RG_V4	SX_V1	FE_V2	FE_V4	FE_V5
0	0,9855	0,9873	0,9878	0,9861	0,9852	0,9850	0,9862	0,9864
1	0,9587	0,9636	0,9651	0,9603	0,9579	0,9571	0,9606	0,9612
2	0,9410	0,9480	0,9501	0,9434	0,9400	0,9388	0,9437	0,9446
3	0,9238	0,9328	0,9355	0,9268	0,9225	0,9210	0,9272	0,9284
4	0,9068	0,9177	0,9210	0,9105	0,9052	0,9034	0,9109	0,9124
5	0,8897	0,9025	0,9064	0,8940	0,8878	0,8858	0,8946	0,8963
6	0,8641	0,8796	0,8843	0,8693	0,8618	0,8593	0,8700	0,8721
7	0,8190	0,8392	0,8453	0,8257	0,8160	0,8127	0,8266	0,8293
8	0,7372	0,7651	0,7737	0,7465	0,7331	0,7286	0,7477	0,7515
9	0,6425	0,6781	0,6892	0,6543	0,6373	0,6317	0,6558	0,6606
10	0,4357	0,4822	0,4971	0,4509	0,4291	0,4220	0,4528	0,4591
11	0,3369	0,3848	0,4004	0,3524	0,3303	0,3232	0,3544	0,3608
12	0,1908	0,2335	0,2481	0,2043	0,1851	0,1791	0,2061	0,2118

Fonte: Autor (2017).

Observe que na Tabela 4.10 são apresentadas as estimativas da função de sobrevivência ajustadas para as covariáveis significativas.

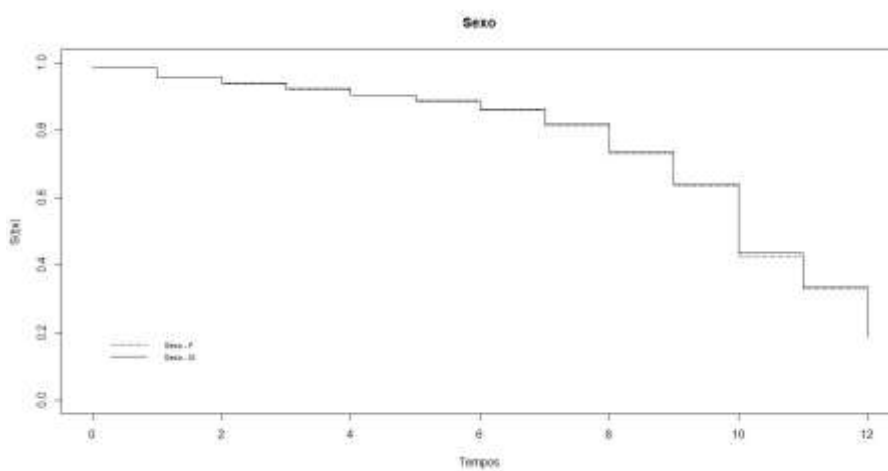
Figura 4.6 – Curvas das Estimativas da Função de Sobrevivência ($S_0(t)$) ajustadas para a covariável Região.



Fonte: Autor (2017).

Note que na Figuras 4.6 a região Norte (var. RG_V2) é a que mais impacta no tempo de resposta satisfatória.

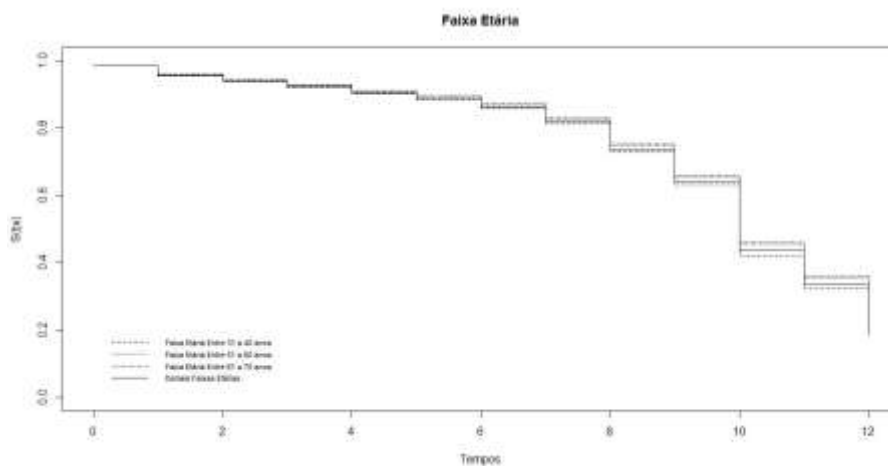
Figura 4.7 – Curvas das Estimativas da Função de Sobrevivência ($S_0(t)$) ajustadas para a covariável Sexo.



Fonte: Autor (2017).

Observe que embora de forma ligeira, os homens aparentam ser mais duros ao considerar a resposta das empresas considerando-as mais insatisfatórias ou não resolvidas que as mulheres.

Figura 4.8 – Curvas das Estimativas da Função de Sobrevivência ($S_0(t)$) ajustadas para a covariável Faixa Etária.



Fonte: Autor (2017).

Note que a Figura 4.8 apresenta as curvas de sobrevivência ajustadas para cada covariável faixa etária sendo que os clientes entre 31 e 40 anos são mais propensos a considerar insatisfatórias as respostas das empresas.

Tabela 4.11 – Estimativas da Função de Taxa de Falha ($\lambda_0(t)$) ajustadas para as covariáveis significativas

Tempo	$\lambda_0(t)$	RG_V1	RG_V2	RG_V4	SX_V1	FE_V2	FE_V4	FE_V5
0	0,0146	0,0128	0,0123	0,0140	0,0149	0,0152	0,0139	0,0137
1	0,0422	0,0371	0,0355	0,0405	0,0430	0,0438	0,0402	0,0395
2	0,0608	0,0534	0,0511	0,0583	0,0619	0,0631	0,0580	0,0570
3	0,0792	0,0696	0,0667	0,0760	0,0807	0,0823	0,0756	0,0742
4	0,0978	0,0859	0,0823	0,0938	0,0996	0,1016	0,0933	0,0917
5	0,1168	0,1026	0,0983	0,1120	0,1190	0,1213	0,1114	0,1095
6	0,1461	0,1283	0,1229	0,1401	0,1487	0,1517	0,1393	0,1369
7	0,1997	0,1754	0,1680	0,1915	0,2034	0,2074	0,1904	0,1872
8	0,3049	0,2677	0,2565	0,2923	0,3105	0,3166	0,2907	0,2857
9	0,4424	0,3885	0,3723	0,4242	0,4505	0,4594	0,4219	0,4146
10	0,8308	0,7295	0,6991	0,7966	0,8460	0,8626	0,7922	0,7786

11	1,0878	0,9551	0,9153	1,0430	1,1077	1,1295	1,0373	1,0194
12	1,6565	1,4544	1,3937	1,5882	1,6867	1,7199	1,5795	1,5523

Fonte: Autor (2017).

Conforme a Tabela 4.11, onde são apresentadas as curvas de taxa de falha sensibilizadas pelas covariáveis, nota-se que as taxas de falhas aumentam com o passar dos dias.

Desta maneira, por exemplo, reclamações feitas por homens, porém oriundas do Centro-Oeste comparadas com a região Sudeste tem-se a seguinte razão de riscos:

$$\frac{\lambda(t | x_{CO;H}) = \exp\{\beta_{RG_V1} + \beta_{SX_V1}\}}{\lambda(t | x_{CO;H}) = \exp\{\beta_{SX_V1}\}} = 0,88$$

o que significa que o risco de uma reclamação oriunda do Centro-Oeste feitas por homens é 13% menor que as feitas por homens nas regiões Sudeste.

4.6 DISCUSSÃO DOS RESULTADOS

A razão de risco, apresentada no Quadro 4.3, significa o número de vezes que o risco da reclamação ser considerada não resolvida pelo cliente (i.e., não foi cumprido o propósito do portal, resposta e soluções rápidas para o consumidor) quando comparado com a média das respostas e soluções do programa.

Quadro 4.3 Resumo dos resultados

Covariável	Classe	Variável Significativa	Razão de Risco*
Região (RG)	CO	RG_V1	0,88
	N	RG_V2	0,84
	S	RG_V4	0,96
Sexo (SX)	M	SX_V1	1,02
Faixa Etária (FE)	entre 31 a 40 anos	FE_V2	1,04
	entre 51 a 60 anos	FE_V4	0,95
	entre 61 a 70 anos	FE_V5	0,94

* vezes quando comparado com as classes não significativas

Fonte: Autor (2017).

Na aplicação prática desse conceito, nota-se que os consumidores das regiões Centro-Oeste, Norte e Sul (CO, N e S respectivamente) possuem um rigor menor para considerar a reclamação atendida dada as respectivas estatísticas de Razão de Risco serem inferiores a 1 (CO = 0,88; N = 0,84 e S = 0,96). Neste mesmo raciocínio, as mulheres são mais rigorosas que os homens (M = 1,02) e que consumidores com mais de 51 anos são mais complacentes com relação as reclamações que os demais.

5 CONSIDERAÇÕES FINAIS

Como resultado desse estudo, nota-se que algumas regiões, o sexo e algumas faixas etárias sensibilizam o entendimento do consumidor em termos de resolução das reclamações feitas no portal Consumidor.com.br. Observa-se que algumas classes das variáveis estudadas são significativas no entendimento da eficiência da plataforma. Desta maneira foi testado e avaliado a adequação do ajuste do modelo de Cox para o tempo de solução de conflitos utilizando a ferramenta virtual que permite a interlocução direta entre consumidores e empresas.

Durante o trabalho, algumas dificuldades se apresentaram e soluções foram tomadas para essas dificuldades. Primeiramente, depois da definição da variável resposta foram identificados diversos registros sem o dado necessário a construção da mesma. Estes dados foram excluídos para o propósito da análise para a construção da variável resposta considerando o tamanho amostral. Adicionalmente, no momento do ajuste do modelo não foram encontrados métodos de seleção de variáveis automáticos já programados no R, como stepwise, para a seleção das variáveis significativas. Para isso foi adotado o método proposto por Collett (1994), entretanto ele aconselha um conhecimento empírico prévio do assunto. Finalmente, na verificação dos pressupostos de proporcionalidade dos riscos, para o uso do modelo de Cox não estendido, o teste estatístico da não suposição foi significativo, contudo este teste é altamente influenciado pelo tamanho amostral. Neste sentido observou-se os métodos gráficos de análise do atendimento da suposição e deixa-se para estudos futuros o impacto do tamanho amostral no resultado desse teste.

Conclui-se que o modelo ajustado identifica algumas das características demográficas que afetam a satisfação do consumidor, no sentido da resolução da reclamação, no processo desenvolvido pelo portal Consumidor.com.br.

5.1 SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestão para trabalhos futuros, algumas técnicas poderiam ser utilizadas para melhorar o ajuste do modelo, como seguem:

- Para etapa 2.1 (Dados):

- Estudar o motivo e possíveis soluções para os dados faltantes na construção da variável resposta;
- Obter e analisar variáveis referentes à reclamação como tipo de produto/serviço (Telecomunicações, Financeiras), tipo de problema (falha na venda, defeito, falha na entrega), entre outras.
- Analisar a possibilidade de amostrar os dados;
- Analisar a evolução das reclamações conforme a disponibilização dos dados no Portal Brasileiro de Dados Abertos;
- Para etapa 4 (Modelagem das Reclamações)
 - Comparar o modelo de Cox com um modelo paramétrico ajustado;
 - Analisar a significância da inclusão de interações duplas e triplas no modelo;
 - Estudar o impacto do tamanho da amostra na obtenção das estimativas do valor p no teste de correlação das variáveis;

REFERÊNCIAS

- BAILAR III, J. C.; MOSTELLER, F. *Medical Uses of Statistics*. 2nd edition, NEJM Book, Boston, 1992.
- COLLETT, D. *Modelling Survival Data in Medical Research*. 1994. Chapman and Hall, London.
- COLOSIMO, Enrico; GIOLO, Sueli. *Análise de Sobrevivência Aplicada – ABE – Projeto Fisher*. 2006.
- COX, D. R. *Regression Models and Life Tables*. ***Journal Royal Statistical Society, B***, n 34, p 187 – 220. 1972.
- KAPLAN, E. L.; MEIER, P. *Nonparametric estimation from incomplete observations*. ***Journal of the American Statistical Association***, n 53, 457-81. 1958.
- RIGONI, Emerson. **Metodologia para implantação da manutenção centrada na confiabilidade**: uma abordagem fundamentada em Sistemas Baseados em Conhecimento e Lógica *Fuzzy*. 2009. F. Dissertação (Doutorado) – Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Santa Catarina. Florianópolis, 2009.
- STRAUHS, Faimara. **Gestão do Conhecimento em Laboratório Acadêmico**: Proposição de Metodologia. 2003. F. Dissertação (Doutorado) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis, 2003.
- WHALEN, Gary. *Economic Review - Federal Reserve Bank of Cleveland*; Cleveland Vol. 27, Iss. 1. 1991.