

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA**

Jonathan Matheus de Freitas Magalhães

**MODELOS DE TARIFAÇÃO: UMA COMPARAÇÃO ENTRE
MÉTODOS DE ESTIMAÇÃO DO PRÊMIO DE RISCO**

BELO HORIZONTE

2022

JONATHAN MATHEUS DE FREITAS MAGALHÃES

MODELOS DE TARIFAÇÃO: UMA COMPARAÇÃO ENTRE
MÉTODOS DE ESTIMAÇÃO DO PRÊMIO DE RISCO

Monografia apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Bacharel em Ciências Atuariais.

Orientadora: Thaís Paiva Galletti

BELO HORIZONTE
2022

JONATHAN MATHEUS DE FREITAS MAGALHÃES

MODELOS DE TARIFAÇÃO: UMA COMPARAÇÃO ENTRE
MÉTODOS DE ESTIMAÇÃO DO PRÊMIO DE RISCO

Monografia apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Bacharel em Ciências Atuariais.

Belo Horizonte, de fevereiro 2022

Banca Examinadora:

Professora Thaís Paiva Galletti
Orientadora
Universidade Federal de Minas Geais

Professora Jussiane Nader Gonçalves
Professora convidada
Universidade Federal de Minas Geais

Agradecimentos

Em primeiro lugar agradeço a Deus por ter iluminado todo o meu caminho até aqui. Quero agradecer também a meus pais e irmão por terem me dado todo o apoio e valores que eu precisava no decorrer dessa caminhada. Aos meus amigos, em especial àqueles que fiz nesse tempo de curso, pois também participaram do processo de aprendizado, pois com todos ajudando foi possível chegar até o final. Aos professores que me ajudaram ao longo de todo curso, fica aqui meus agradecimentos, em especial à minha orientadora Thaís que deu todo o suporte que eu precisava para concluir esse trabalho e também à Jussiane que também ajudou em pontos importantes desse trabalho.

Resumo

O Brasil tem um número impressionante de veículos em sua frota. Consequentemente, é esperado uma grande ocorrência de sinistros tais como acidentes, roubo/furto, dentre outros. Por isso, é de suma importância realizar o seguro do veículo, que consiste em transferir o risco de sinistro para uma seguradora mediante o pagamento de um prêmio. Esse fato torna importante o cálculo de uma estimativa para o prêmio usando as características do bem segurado e do motorista, de modo que esse prêmio seja mais realista, o que reflete diretamente na situação econômica-financeira e atuarial da seguradora. Sendo assim, o presente trabalho buscou estimar o prêmio de risco para a Região Sudeste usando dados do segundo semestre de 2019 disponibilizados pela Superintendência de Seguros Privados. Para isso, foram testados e ajustados modelos para a frequência usando as distribuições de Poisson, Binomial Negativa e Log-Normal, e para a severidade usando as distribuições Gama e Log-Normal. As distribuições que apresentaram ajustes mais adequados foram Log-Normal para frequência e Gama para a severidade. Com as relatividades desses modelos, foi estimado o Prêmio de Risco usando o método multiplicativo, para cada célula tarifária de acordo com o tipo de cobertura. Adicionado a isso, foi ajustado um modelo usando a distribuição de mistura Tweedie para realizar uma comparação com Prêmio de Risco estimado através do método multiplicativo.

Abstract

Brazil has an impressive number of vehicles in its fleet. Consequently, a large occurrence of claims such as accidents, theft, among others, is expected. Therefore, carrying out vehicle insurance, which consists of transferring the risk of a claim to an insurer given the payment of the premium. This makes it important to evaluate an estimate for the premium using the characteristics of the insured vehicle and the driver. Therefore, the present work sought to estimate the semester premium for the Southeast Region of Brazil using data from the second semester of 2019 made available by the “Superintendência de Seguros Privados”, the national agency responsible for surveillance of the private insurance market. For this, models were tested and adjusted for frequency using Poisson, Negative Binomial and Log-Normal distributions, and for severity using Gamma and Log-Normal distributions. The distributions that presented the most adequate fit were Log-Normal and Gamma for severity. With the relativities of these models, the Risk Premium was estimated using the multiplicative method, for each tariff cell according to the type of coverage. In addition to that, it was also adjusted a model with the Tweedie mixture distribution to compare the results with the Risk Premium obtained from the multiplicative model.

Sumário

1	INTRODUÇÃO	10
2	REVISÃO DA LITERATURA	12
3	METODOLOGIA	14
3.1	Dados	14
3.2	Prêmios	19
3.3	Modelos Lineares Generalizados	19
3.3.1	Família Exponencial	19
3.3.2	Componentes de um MLG	20
3.3.3	Métodos de estimação para um MLG	21
3.3.4	Avaliação do ajuste de um MLG	22
3.4	Células Tarifárias	23
3.5	Modelo Tweedie	24
4	RESULTADOS	25
4.1	Análise Descritiva	25
4.1.1	Sinistros da cobertura roubo/furto	27
4.1.2	Sinistros da cobertura colisão parcial	29
4.1.3	Sinistros da cobertura colisão perda total	30
4.1.4	Sinistros da cobertura incêndios	32
4.1.5	Sinistros de outras coberturas	34
4.2	Ajuste e interpretação dos modelos para Frequência e Severidade	36
4.3	Ajuste e interpretação dos modelos Tweedie	42
4.4	Prêmio de Risco	44
5	CONCLUSÃO	48
6	REFERÊNCIAS	49
	ANEXO A – MODELOS PARA FREQUÊNCIA	51
	ANEXO B – MODELOS PARA SEVERIDADE	53
	ANEXO C – MODELOS TWEEDIE	55

Lista de tabelas

Tabela 3.1 – Descrição das variáveis originais do banco de dados	15
Tabela 3.2 – Análise preliminar da quantidade de sinistros por código tarifário . . .	16
Tabela 3.3 – Descrição das variáveis que serão usadas na modelagem	18
Tabela 3.4 – Parâmetros canônicos e de ruído de algumas distribuições da família exponencial	20
Tabela 4.1 – Regiões e suas respectivas frequências de sinistros por tipo de cobertura	26
Tabela 4.2 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Incêndios	38
Tabela 4.3 – Relatividades para Frequência da cobertura Incêndios	39
Tabela 4.4 – Resultado do ajuste do modelo Gama para Severidade da cobertura Incêndios	40
Tabela 4.5 – Relatividades para Severidade da cobertura Incêndios	41
Tabela 4.6 – Resultado do ajuste do modelo Tweedie para cobertura Incêndios . . .	42
Tabela 4.7 – Relatividades do modelo Tweedie para cobertura Incêndios	43
Tabela 4.8 – Prêmio de Risco estimado para cobertura Incêndios usando modelo multiplicativo	44
Tabela 4.9 – Prêmio de Risco estimado para cobertura Incêndios usando modelo Tweedie	45
Tabela A.1 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Roubo/Furto	51
Tabela A.2 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Colisão Parcial	51
Tabela A.3 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Colisão Perda Total	52
Tabela A.4 – Resultado do ajuste do modelo Log-Normal para Frequência de outras coberturas	52
Tabela B.1 – Resultado do ajuste do modelo Gama para Severidade da cobertura Roubo/Furto	53
Tabela B.2 – Resultado do ajuste do modelo Gama para Severidade da cobertura Colisão Parcial	53
Tabela B.3 – Resultado do ajuste do modelo Gama para Severidade da cobertura Colisão Perda Total	54
Tabela B.4 – Resultado do ajuste do modelo Gama para Severidade de outras coberturas	54
Tabela C.1 – Resultado do ajuste do modelo Tweedie para cobertura Roubo/Furto .	55
Tabela C.2 – Resultado do ajuste do modelo Tweedie para cobertura Colisão Parcial	55

Tabela C.3–Resultado do ajuste do modelo Tweedie para cobertura Colisão Perda	
Total	56
Tabela C.4–Resultado do ajuste do modelo Tweedie para outras coberturas	56

Lista de figuras

Figura 3.1 – Distribuição das variáveis exposição e quantidade de sinistros da cobertura roubo/furto entre 1990 e 2020.	17
Figura 4.1 – Gráficos de exposição e quantidade de sinistros (de acordo com o tipo de cobertura) por região	25
Figura 4.2 – Frequência 1 por cód. tarif.	27
Figura 4.3 – Frequência 1 por ano modelo	27
Figura 4.4 – Frequência 1 por idade	27
Figura 4.5 – Frequência 1 por região e sexo	27
Figura 4.6 – Severidade 1 por cód. tarif.	28
Figura 4.7 – Severidade 1 por ano modelo	28
Figura 4.8 – Severidade 1 por idade	28
Figura 4.9 – Severidade 1 por região e sexo	28
Figura 4.10–Frequência 2 por cód. tarif.	29
Figura 4.11–Frequência 2 por ano modelo	29
Figura 4.12–Frequência 2 por idade	29
Figura 4.13–Frequência 2 por região e sexo	29
Figura 4.14–Severidade 2 por cód. tarif.	30
Figura 4.15–Severidade 2 por ano modelo	30
Figura 4.16–Severidade 2 por idade	30
Figura 4.17–Severidade 2 por região e sexo	30
Figura 4.18–Frequência 3 por cód. tarif.	31
Figura 4.19–Frequência 3 por ano modelo	31
Figura 4.20–Frequência 3 por idade	31
Figura 4.21–Frequência 3 por região e sexo	31
Figura 4.22–Severidade 3 por cód. tarif.	31
Figura 4.23–Severidade 3 por ano modelo	31
Figura 4.24–Severidade 3 por idade	32
Figura 4.25–Severidade 3 por região e sexo	32
Figura 4.26–Frequência 4 por cód. tarif.	33
Figura 4.27–Frequência 4 por ano modelo	33
Figura 4.28–Frequência 4 por idade	33
Figura 4.29–Frequência 4 por região e sexo	33

Figura 4.30–Severidade 4 por cód. tarif.	33
Figura 4.31–Severidade 4 por ano modelo	33
Figura 4.32–Severidade 4 por idade	34
Figura 4.33–Severidade 4 por região e sexo	34
Figura 4.34–Frequência 9 por cód. tarif.	34
Figura 4.35–Frequência 9 por ano modelo	34
Figura 4.36–Frequência 9 por idade	35
Figura 4.37–Frequência 9 por região e sexo	35
Figura 4.38–Severidade 9 por cód. tarif.	35
Figura 4.39–Severidade 9 por ano modelo	35
Figura 4.40–Severidade 9 por idade	36
Figura 4.41–Severidade 9 por região e sexo	36
Figura 4.42–Hnp para Frequência da cobertura Incêndios	39
Figura 4.43–Hnp para Severidade da cobertura Incêndios	41
Figura 4.44–Gráficos com Prêmio de Risco estimado para cobertura Incêndios usando modelo multiplicativo	45
Figura 4.45–Gráficos com Prêmio de Risco estimado para cobertura Incêndios usando modelo Tweedie	46

1 Introdução

Vivemos em um país que, segundo o Ministério da Infraestrutura, em outubro de 2021, possuía mais de 110 milhões de veículos em sua frota¹. Isso quer dizer que, para cada dois brasileiros, temos em torno de um veículo (baseado na projeção que o IBGE expôs no dia 14 de junho de 2021 de que o Brasil possui mais de 213 milhões de pessoas na população)². Na região Sudeste, região escolhida para o estudo, temos uma frota de mais de 53 milhões de veículos, sendo que, desta frota, cerca de 32 milhões são automóveis e 9 milhões são motocicletas.

Devido a esse grande número de veículos, espera-se que a ocorrência de sinistros (acidentes, furtos, dentre outros) também seja grande, existindo também a possibilidade de sinistros com óbitos. Dados do Departamento de Informática do Sistema Único de Saúde do Brasil (DATASUS) mostram que, em 2019 tivemos 31.945 óbitos no Brasil cuja a causa primária da morte foi um acidente de transporte³. Desses óbitos, 33,32% aconteceram na região Sudeste.

Esses números expressivos de veículos na frota brasileira, e também de óbitos relacionados com o uso dos meios de transportes, revelam a importância do setor de seguros de autos no país. O órgão que controla e fiscaliza esse setor é a Superintendência de Seguros Privados (SUSEP). Tal órgão é uma “entidade autárquica dotada de personalidade jurídica de Direito Público, com autonomia administrativa e financeira”. Dados divulgados por essa instituição na “Síntese Mensal de Outubro de 2021” confirmam a importância do setor citado anteriormente. Nessa síntese, a SUSEP mostra que o setor de seguro auto arrecadou R\$ 30,80 bilhões em prêmios até outubro, o que representou um aumento de 7% em relação ao mesmo período de 2020. Além disso, o valor arrecadado de prêmios de seguro auto representou 41,8% do valor total arrecadado com prêmio com seguro de danos (seguro auto é uma subdivisão de seguro de danos).

Ainda dentro desta linha de raciocínio, dados da Confederação Nacional de Seguros (CNseg) mostram que a arrecadação de prêmios de seguros de ramos elementares (ramo no qual o seguro auto está incluído), de janeiro de 2021 até outubro de 2021, representou 1% do PIB nominal do Brasil. Ademais, a penetração com arrecadação de prêmios de seguros de forma geral no PIB nominal de 2020 foi de 6,7%⁴.

¹ Ministério da Infraestrutura. **Frota de Veículos - 2021**. Disponível em <https://www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-denatran/frota-de-veiculos-2021> Acesso em: 07 de janeiro de 2022

² IBGE. **Projeção da População**. Disponível em <https://www.ibge.gov.br/apps/populacao/projecao/index.html> Acesso em: 14 de junho de 2021

³ DATASUS. **Óbitos por causas externas - Brasil**. Disponível em <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/ext10uf.def> Acesso em: 20 de junho de 2021

⁴ CNseg. **Estatísticas do Setor Segurador**. Disponível em <https://cnseg.org.br/analises-e-estatisticas/panorama-estatistico-8A8AA8A37A39FA6A017AC49227594A3A.html>

Ao longo dos parágrafos anteriores, foram apresentados diversos dados estatísticos e financeiros que servem para ressaltar a importância desse trabalho, cujo tema será **Modelos de Tarifação: Uma comparação entre métodos de estimação do prêmio de risco**.

Para precificarmos seguros do ramo não vida (ou ramos elementares), utilizamos de modelos que fazem uso de algumas variáveis que representam características observáveis que afetam o risco de acidentes e o valor das indenizações, tais como: as características do motorista, localidade, características do objeto segurado. Tais modelos foram introduzidos na precificação por Jewell (1974). Para auxiliar no correto dimensionamento do efeito de cada uma das variáveis, é necessário fazer uma análise dos dados sobre frequência e severidade dos sinistros registrados.

Sendo assim, o objetivo deste trabalho é analisar os efeitos das características do segurado e do veículo sobre a frequência e a severidade dos sinistros na região Sudeste do Brasil, com base em informações para o segundo semestre de 2019, retiradas do sistema AUTOSEG⁵ da SUSEP. Além disso, esse trabalho busca comparar alguns métodos de tarifação de seguros.

No Capítulo 2 será feita uma breve revisão dos trabalhos já realizados nessa temática. No Capítulo 3 serão apresentadas as variáveis que temos na base de dados, além de conceitos importantes que serão usados nesse trabalho. No Capítulo 4 será realizada uma análise detalhada dos dados, além de interpretação dos ajustes e Prêmios de Riscos construídos com base na célula tarifária estimada. E, por fim, nos Capítulos 5 e 6 apresentamos, respectivamente, a conclusão desse estudo e as referências que foram utilizadas.

Acesso em: 7 de janeiro de 2022

⁵ AUTOSEG. Disponível em <http://www2.susep.gov.br/menuestatistica/Autoseg/menu1.aspx>
Acesso em: 29 de junho de 2021

2 Revisão da Literatura

Souza et al. (2000) define seguro como: “contrato pelo qual uma das partes se obriga, mediante cobrança de prêmio, a indenizar outra pela ocorrência de determinados eventos ou por eventuais prejuízos”. Os eventos citados anteriormente são denominados sinistros, definido por Morgado (2004) como “o evento responsável por algum dano sobre o bem segurado, ou mesmo qualquer dano provocado ao bem de terceiros”.

Já o prêmio é definido pela Escola Nacional de Seguros (2011) como o valor pago pelo segurado à seguradora pela transferência do risco, sendo esse valor definido no contrato de seguro. Ferreira (2002) afirma que existem três tipos de prêmios no processo de precificação:

- 1) Prêmio de Risco: cobre o risco médio;
- 2) Prêmio Puro: é o prêmio de risco acrescido de um carregamento estatístico;
- 3) Prêmio Comercial: é o prêmio puro acrescido do carregamento das demais despesas da seguradora, incluindo uma margem de lucro.

Para a estimação do prêmio do seguro, Ohlsson e Johansson (2010) afirmam que é necessária a utilização de métodos estatísticos, devido à aleatoriedade da taxa de acidentes e do valor da perda. E, dentro dessa linha de raciocínio, Turkman e Silva (2000) explicam que, para quem utiliza desses métodos, é necessário o conhecimento da metodologia e forma de aplicação dos Modelos Lineares Generalizados (MLG), pois a metodologia desses modelos unifica procedimentos estatísticos usados na prática e faz o uso do método da máxima verossimilhança, que é um elemento central na teoria da inferência. Desta forma, um dos métodos mais utilizados na tarifação de seguros é o MLG.

Os MLG's foram introduzidos por Nelder e Wedderburn (1972), que fizeram uma síntese de modelos lineares ou não lineares apresentados antes do trabalho feito por eles. Tais modelos eram construídos usando distribuições que pertenciam a uma família de distribuições denominada Família Exponencial, que será explicada com mais detalhes na Seção 3.3. Silvério (2015) afirma que o trabalho de McCullagh e Nelder (1989) consolidou e formalizou a classe de MLG, incluindo o modelo linear clássico e modelos que usa as funções de ligação logit e probit, e o linear clássico. Esses modelos realizam a estimação de coeficientes, que em tarifação, são utilizados para a estimação do Prêmio de Risco.

Morgado (2004) mostra que foi um processo longo para que a precificação dos seguros chegasse ao nível técnico que se encontra hoje. Em seu trabalho, o autor aponta o estudo de Mowbray (1914), um dos primeiros trabalhos em que foi utilizado método estatístico para determinar o valor de um prêmio, e explica que Mowbray pressupôs a distribuição binomial dos sinistros, e provou que, com a experiência individual de um ano, o prêmio puro estimado poderia ser generalizado para o ano seguinte. Entretanto, o trabalho de Mowbray tinha limitação de que o número de sinistros não poderia passar de um

determinado valor. Em seguida, Morgado apresenta uma série de estudos que procuraram propor soluções para a precificação de um seguro e afirma que, somente depois do trabalho de Jewell (1974) que “mostrou a relação entre a família de distribuições exponenciais e a teoria da credibilidade”, que o MLG foi introduzido na precificação.

Com o objetivo de precificar e utilizando dados de uma seguradora denominada Wasa, presentes no livro de Ohlsson e Johansson (2010), Silvério (2015) iniciou sua modelagem ajustando um modelo de Poisson para a frequência de sinistros e quatro modelos para a severidade, e obteve como melhor ajuste os modelos Log-Poisson e Log-Normal Inversa para frequência e severidade, respectivamente. E, por fim, realizou a modelagem do prêmio que consiste em multiplicar as relatividades correspondentes aos modelos estimados para frequência e severidade.

Na literatura, existem alguns trabalhos que objetivam fazer comparações entre métodos de tarifação. Silva e Afonso (2015), usando dados do Brasil, fizeram um estudo comparando quatro métodos de precificação: Modelagem do Prêmio Puro por sinistros historicamente agregados, Modelagem do Prêmio Puro por sinistros previstos agregados, Modelos Clássicos Lineares e Modelos Lineares Generalizados. Os autores concluíram que o MLG e a modelagem de Prêmio Puro histórico apresentaram melhores resultados, indicando que para portfólios médios, o MLG se mostra como uma boa alternativa de precificação.

Outro estudo comparativo foi o de Xacur & Garrido (2015) que fizeram uma comparação entre o ajuste da modelagem de frequência e severidade separados e o ajuste do modelo Tweedie (Tweedie, 1984), que é uma modelagem conjunta da frequência e severidade, do ponto de vista do: número de parâmetros, variabilidade do parâmetro de dispersão, otimização das estimativas dos coeficientes, e perda de informação. Os autores explicam que a distribuição Tweedie é uma distribuição de mistura que pertence à família exponencial, e que essa distribuição tem uma massa de probabilidade positiva no zero, representando os casos sem ocorrência de sinistros, e um componente contínuo para valores positivos das indenizações. Na conclusão do estudo, Xacur & Garrido indicam que, apesar das limitações, o modelo Tweedie é uma boa alternativa de precificação.

Por fim, um dos trabalhos mais recentes a respeito de tarifação de seguro auto no Brasil é exposto por Farias e de Jesus (2020). O estudo buscou desenvolver uma tabela tarifária para a região Nordeste do Brasil, por meio de MLG, utilizando dados do AUTOSEG referentes ao 1º semestre de 2018. A partir desses dados, ajustaram a frequência usando modelo Poisson e a severidade usando modelo Log-Normal e estimaram o prêmio puro da forma apresentada por Silvério (2015). A conclusão que eles chegaram é a de que foi possível construir uma tabela tarifária para a região Nordeste, mas que era necessário o estudo de outras distribuições e uma melhor avaliação dos ajustes dos modelos. Tomando como base o trabalho de Farias e de Jesus (2020), o presente estudo buscará explorar o que foi concluído pelos autores.

3 Metodologia

Ao longo dos próximos capítulos, quando estivermos nos referindo a frequência e severidade das indenizações, estaremos usando as seguintes definições:

- 1) **Frequência:** Quantidade de sinistros dividido pelo número de expostos;
- 2) **Severidade:** Total de indenizações dividido pela quantidade de sinistros.

Nesse Capítulo faremos uma descrição do banco de dados na Seção 3.1. Ademais, nas seções 3.2, 3.3, 3.4, 3.5 faremos algumas definições sobre prêmios, MLG, células tarifárias e modelo tweedie, respectivamente.

3.1 Dados

Os dados utilizados no presente estudo foram extraídos do AUTOSEG, que é um sistema da SUSEP que disponibiliza estatísticas e dados dos automóveis segurados. Os dados selecionados são referentes ao segundo semestre de 2019 para a região Sudeste. Para realizar as análises descritivas e modelagem, o programa estatístico R⁶ foi utilizado.

Na Tabela 3.1, serão apresentadas as variáveis que compõem esse banco de dados e suas respectivas descrições.

⁶ R. Disponível em <https://www.r-project.org/> Acesso em: 20 de dezembro de 2021

Tabela 3.1 – Descrição das variáveis originais do banco de dados

Variável	Descrição
COD TARIF	Código Tarifário 1 = Passeio nacional 2 = Passeio Importado 3 = Pick-up (nacional e importado) 4 = Veículo de Carga (nacional e importado) 5 = Motocicleta (nacional e importado) 6 = Ônibus (nacional e importado) 7 = Utilitários (nacional e importado) 9 = Outros
REGIAO	9 = SP: Vale do Paraíba e Ribeira 10 = SP: Litoral Norte e Baixada Santista 11 = SP: Met. de São Paulo 12 = SP: Grande Campinas 13 = SP: Ribeirão Preto e Demais Mun. de Campinas 14 = MG: Triângulo mineiro 15 = MG: Sul 16 = MG: Met.BH-Centro Oeste-Zona Mata-C. Vertentes 17 = MG: Vale do Aço-Norte-Vale Jequitinhonha 18 = RJ: Met. do Rio de Janeiro 19 = RJ: Interior 20 = ES: Espírito Santo
ANO MODELO	Ano do modelo do veículo (temos de 1990 até 2021)
SEXO	Sexo do Condutor M = Masculino F = Feminino
IDADE	Idade do Condutor 0 = Não informada 1 = Entre 18 e 25 anos 2 = Entre 26 e 35 anos 3 = Entre 36 e 45 anos 4 = Entre 46 e 55 anos 5 = Maior que 55 anos
EXPOSICAO1	Quantidade de veículos expostos
PREMIO1	Prêmio médio das apólices
IS MEDIA	Média das Importâncias Seguradas das apólices incluídas no grupamento definido pela chave escolhida, ponderada pela exposição de cada uma delas
FREQ SIN1	Quantidade de sinistros da cobertura roubo/furto
INDENIZ1	Total de indenizações de sinistros da cobertura roubo/furto
FREQ SIN2	Quantidade de sinistros da cobertura colisão parcial
INDENIZ2	Total de indenizações de sinistros da cobertura colisão parcial
FREQ SIN3	Quantidade de sinistros da cobertura colisão perda total
INDENIZ3	Total de indenizações de sinistros da cobertura colisão perda total
FREQ SIN4	Quantidade de sinistros da cobertura incêndio
INDENIZ4	Total de indenizações de sinistros da cobertura incêndio

FREQ SIN9	Quantidade de sinistros de outras coberturas, como assistência 24 hs, etc
INDENIZ9	Total de indenizações de sinistros de outras coberturas, como assistência 24 hs, etc

Como pode ser percebido na Tabela 3.1, as variáveis COD TARIF, REGIAO e ANO MODELO apresentam muitas categorias. Desta forma, foi realizado um breve estudo para verificar qual a melhor maneira de se trabalhar com elas.

Para analisar COD TARIF, verificamos o número de vezes que cada código tarifário aparece no banco de dados, em relação ao total. Além disso, agregamos as frequências de cada tipo de sinistro e depois verificamos a frequência que cada código tarifário apresentou em relação à frequência total. O mesmo procedimento foi aplicado para o valor da indenização e para a exposição. Na Tabela 3.2 abaixo, temos os resultados encontrados.

Tabela 3.2 – Análise preliminar da quantidade de sinistros por código tarifário

Código Tarifário	Porcentagem dos dados	Frequência	Indenização	Exposição
1	56,00%	0,77	0,60	0,74
2	12,00%	0,04	0,07	0,04
3	22,00%	0,17	0,26	0,17
4	2,00%	0,01	0,02	0,00
5	7,00%	0,02	0,03	0,05
6	0,00%	0,00	0,00	0,00
7	0,00%	0,00	0,00	0,00
9	1,00%	0,00	0,02	0,00

Pela Tabela 3.2, vemos que a maior parte dos dados se encontram nos códigos 1, 2, 3 e 5, logo, convêm que trabalhemos apenas com essas categorias, já que as demais não são significativas. Conseqüentemente, iremos trabalhar apenas com os códigos tarifários 1 (Passeio Nacional), 2 (Passeio Importado), 3 (Pick-up) e 5 (Motocicleta).

Quanto ao ANO MODELO, como podemos ver nos gráficos da Figura 3.1, tanto para exposição semestral quanto para quantidade de sinistros semestral de roubo/furto, temos uma concentração da maior parte dos dados de 2010 até 2020, com os dois gráficos apresentando comportamentos semelhantes. Sendo assim, optou-se por agrupar os anos de 1990 até 2010, e o restante agrupar de forma que as categorias fiquem com intervalos de anos parecidos. A classificação final se encontra na Tabela 3.3.

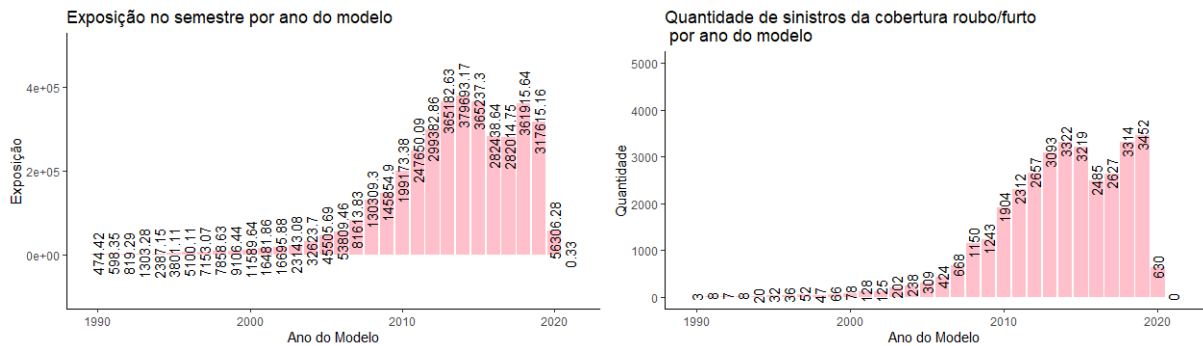


Figura 3.1 – Distribuição das variáveis exposição e quantidade de sinistros da cobertura roubo/furto entre 1990 e 2020.

Já para a variável REGIAO, pretendemos utilizar uma das maneiras utilizadas no mercado para regionalizar. Para isso, um profissional que trabalha em uma empresa de proteção veicular foi consultado sobre como é feita a regionalização de cada estado nessa companhia. A resposta obtida foi a de que nessa rede de proteção veicular (com cobertura nacional) cada estado é separado por Região Metropolitana e Interior. Desta forma, na Tabela 3.3, são apresentadas as variáveis que serão utilizadas para a modelagem e suas respectivas descrições/categorias.

Vale destacar que, nas descrições e títulos de tabelas e figuras, os tipos de cobertura estarão codificados da seguinte forma:

- 1 - Sinistros da cobertura Roubo/Furto
- 2 - Sinistros da cobertura Colisão Parcial
- 3 - Sinistros da cobertura Colisão Perda Total
- 4 - Sinistros da cobertura Incêndios
- 9 - Sinistros de outras coberturas

Tabela 3.3 – Descrição das variáveis que serão usadas na modelagem

Variável	Descrição
COD TARIF	1 = Passeio nacional 2 = Passeio Importado 3 = Pick-up (nacional e importado) 5 = Motocicleta (nacional e importado)
REGIAO	1 = SP: Met. de São Paulo 2 = SP: Interior 3 = MG: Met. de Belo Horizonte 4 = MG: Interior 5 = RJ: Met. do Rio de Janeiro 6 = RJ: Interior 7 = ES: Espírito Santo
ANO MODELO	1 = 1990-2010 2 = 2011-2013 3 = 2014-2017 4 = 2018-2021
SEXO	Sexo do Condutor M = Masculino F = Feminino
IDADE	Idade do Condutor 0 = Não informada 1 = Entre 18 e 25 anos 2 = Entre 26 e 35 anos 3 = Entre 36 e 45 anos 4 = Entre 46 e 55 anos 5 = Maior que 55 anos
EXPOSICAO1	Quantidade de veículos expostos
PREMIO1	Prêmio médio das apólices
IS MEDIA	Média das Importâncias Seguradas das apólices incluídas no grupamento definido pela chave escolhida, ponderada pela exposição de cada uma delas
FREQ SIN1	Quantidade de sinistros da cobertura roubo/furto
INDENIZ1	Total de indenizações de sinistros da cobertura roubo/furto
FREQ SIN2	Quantidade de sinistros da cobertura colisão parcial
INDENIZ2	Total de indenizações de sinistros da cobertura colisão parcial
FREQ SIN3	Quantidade de sinistros da cobertura colisão perda total
INDENIZ3	Total de indenizações de sinistros da cobertura colisão perda total
FREQ SIN4	Quantidade de sinistros da cobertura incêndio
INDENIZ4	Total de indenizações de sinistros da cobertura incêndio
FREQ SIN9	Quantidade de sinistros de outras coberturas, como assistência 24 hs, etc
INDENIZ9	Total de indenizações de sinistros de outras coberturas, como assistência 24 hs, etc

3.2 Prêmios

Na revisão da literatura, foram apresentados os conceitos de cada tipo de prêmio. Agora, iremos defini-los como fórmulas baseado em Ferreira (2002).

O prêmio de risco pode ser formulado da seguinte maneira:

$$\text{Prêmio de Risco} = PR = \text{frequência} \times \text{severidade}$$

O carregamento estatístico, que será acrescido ao prêmio de risco para formar o prêmio puro. É definido como θ . Sendo assim:

$$\text{Prêmio Puro} = PP = PR (1 + \theta)$$

Por fim, definimos α como o carregamento das demais despesas administrativas da seguradora (incluindo margem de lucro), que será aplicado como um percentual do prêmio puro para que cheguemos ao valor do prêmio comercial, formulado da seguinte forma:

$$\text{Prêmio Comercial} = PC = \frac{PP}{1 - \alpha} = \frac{PR (1 + \theta)}{1 - \alpha}$$

Vale destacar que o prêmio que será tabelado através do MLG é o prêmio de risco.

3.3 Modelos Lineares Generalizados

3.3.1 Família Exponencial

Dobson e Barnett (2018) afirmam que a família exponencial pode ser definida como o conjunto de distribuições cuja função pode ser escrita com a seguinte formulação:

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

São exemplo de distribuições que fazem parte dessa família: binomial, binomial negativa, poisson, exponencial, gama, normal, dentre outras.

Dentre as distribuições pertencentes à família exponencial, teremos casos em que $a(y) = y$. Quando esse fato acontece, o $b(\theta)$ será denominado como parâmetro canônico da distribuição. Para exemplificar a descrição realizada acima, vamos considerar Y uma variável aleatória com distribuição de Poisson com parâmetro θ . A função massa de probabilidade dessa distribuição pode ser escrita da seguinte forma:

$$f(y; \theta) = \frac{\theta^y e^{-\theta}}{y!} = \frac{e^{-\theta}}{y!} \exp[y \ln(\theta)] = \exp[y \ln(\theta) - \ln(\theta) - \ln(y!)]$$

Podemos notar que: $a(y) = y$; $b(\theta) = \ln(\theta)$; $c(\theta) = -\ln(\theta)$ e $d(y) = -\ln(y!)$. Portanto, como $a(y) = y$, $b(\theta) = \ln(\theta)$ será o parâmetro canônico da distribuição de Poisson.

Quando estamos lidando com distribuições com dois parâmetros e que possui a formulação canônica da família exponencial, a sua função pode ser reorganizada da seguinte forma:

$$f(y; \theta, \phi) = \exp \left[\frac{yb_*(\theta) - c_*(\theta)}{\phi} + d(y, \phi) \right]$$

onde $\phi > 0$ é o parâmetro de ruído da distribuição, que pode ser entendido como uma medida que controla a dispersão da distribuição. Ademais, $b_*(\theta) = \phi b(\theta)$ e $c_*(\theta) = -\phi c(\theta)$. Na Tabela 3.4, temos um resumo dos parâmetros canônicos e parâmetro de ruído de algumas distribuições.

Tabela 3.4 – Parâmetros canônicos e de ruído de algumas distribuições da família exponencial

Distribuição	$b(\theta)$	ϕ
Poisson	$\ln(\theta)$	1
Binomial	$\ln\left(\frac{\theta}{1-\theta}\right)$	1
Binomial Negativa	$\ln(\theta)$	1
Exponencial	$-\frac{1}{\theta}$	1
Gama	$-\frac{\alpha}{\theta}$	$\frac{1}{\alpha}$
Normal	$\frac{\theta}{\sigma^2}$	σ^2

Como indicado pela Tabela 3.4, as distribuições Binomial e Binomial Negativa, apesar de serem bi-paramétricas, apresentam parâmetro de ruído igual a 1. Isso pode ser explicado pelo fato de que o número de ensaios (para binomial) e o número de falhas (para binomial negativa) normalmente são conhecidos em aplicações que envolvem essas distribuições.

3.3.2 Componentes de um MLG

São três os componentes que fazem parte de um MLG:

1) Componente Aleatório: uma distribuição de probabilidade que pertence à Família Exponencial;

2) Componente Sistemático: preditor linear, η_i , que é uma combinação linear dos regressores usando os coeficientes em β que incorpora informações das variáveis independentes (covariáveis) no modelo;

$$\eta_i = x_i^T \beta$$

3) Função de Ligação: é a função que relaciona o valor esperado da variável resposta, μ_i , ao preditor linear. É uma função $g(\cdot)$ tal que:

$$E(Y_i) = \mu_i = g(X_i^T \beta)$$

Vale ressaltar que é preferível que a imagem da função de ligação esteja definida no intervalo em que os valores possíveis para μ_i estão definidos. Portanto, em alguns casos,

algumas funções de ligação não canônicas serão mais interessantes de serem usadas do que a canônica. Uma outra justificativa para utilização da versão não canônica é o fato de que algumas tornam a interpretação do resultado do modelo mais fácil.

3.3.3 Métodos de estimação para um MLG

O método mais usado para estimar os parâmetros de um MLG é o Método de Máxima Verossimilhança (EMV). A seguir será descrito o processo para a obtenção do EMV e será utilizado um exemplo para deixar claro o procedimento.

Seja $f(y; \theta)$ uma função massa de probabilidade (f.m.p) ou função de distribuição de probabilidade (f.d.p), dada por:

$$u(\theta) = \prod_{i=1}^n f(y_i; \theta)$$

E, por fim, consideraremos que $L(y; \theta) = \ln(u(\theta))$.

O EMV de θ é definido por Bolfarine e Sandoval (2000) como o elemento $\hat{\theta}$ que maximiza a função de verossimilhança, e tal elemento é dado pelo seguinte sistema de equações:

$$\frac{dL(y; \theta)}{d\theta_i} = 0$$

É importante ressaltar que o EMV pode: não ser único; não existir, caso a função não tenha máximo; não ser obtido de forma analítica, ou seja, não ter forma fechada.

Para exemplificarmos o processo descrito acima, vamos considerar Y como uma variável aleatória com distribuição Poisson(λ). Temos que:

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}, y = 0, 1, 2, \dots$$

$$u(\theta) = \prod_{i=1}^n P(Y = y_i) = \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!}$$

Tomando o logaritmo dos dois lados da equação temos:

$$L(y_i; \theta) = \ln \left(\frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!} \right) = \frac{1}{\prod_{i=1}^n y_i!} + \ln(\lambda) \sum_{i=1}^n y_i - n\lambda$$

e derivando em relação a λ , segue que

$$\frac{dL(y; \theta)}{d\theta} = \frac{1}{\lambda} \sum_{i=1}^n y_i - n$$

Igualando a zero, temos:

$$\begin{aligned} \frac{1}{\lambda} \sum_{i=1}^n y_i - n &= 0 \\ \frac{1}{\lambda} \sum_{i=1}^n y_i &= n \end{aligned}$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n y_i$$

Bolfarine e Sandoval (2000) define $\frac{\sum_{i=1}^n y_i}{n}$ como \bar{y} . Logo, $\hat{\lambda} = \bar{y}$. Portanto, \bar{y} é o estimador de máxima verossimilhança para $\hat{\lambda}$.

Temos outros métodos de estimação de parâmetros do MLG, entretanto, como eles não serão utilizados no processo de modelagem, não serão citados nesse trabalho. Em casos de estrutura de regressão, a estimação deve ser feita numericamente, realizado usando a função `glm` do R.

3.3.4 Avaliação do ajuste de um MLG

Existem alguns métodos para realizar a avaliação do ajuste de um modelo, entretanto, nesse trabalho serão apresentados dois métodos: *Deviance* e gráficos *half-normal* de um modelo.

Antes de falar sobre o método da *Deviance*, temos que definir Modelo Sob Pesquisa como o modelo que estamos avaliando, e Modelo Saturado que é o modelo cujo número de parâmetros é igual ao número de observações, Nesse tipo de modelo, a média μ_i é específica para cada observação Y_i , portanto, iremos estimar $\hat{\mu}_i = y_i$.

Sendo assim, a *Deviance* é um método de avaliação do ajuste de modelo que compara o ajuste do Modelo Sob Pesquisa com o ajuste do Modelo Saturado. Sua formulação pode ser descrita como duas vezes a diferença entre a log-verossimilhança do Modelo Saturado e a log-verossimilhança do Modelo Sob Pesquisa. Mais detalhes sobre essa formulação podem ser encontrados no livro dos autores Dobson e Barnett (2018).

A *Deviance* sempre assume valores reais positivos, ou seja, $D > 0$. Além disso, o Teorema de Wilks (que é explicado com detalhes no trabalho de Bastos (2014)) garante que a distribuição assintótica da *Deviance* é Qui-Quadrado com $(n-p)$ graus de liberdade, sendo n o tamanho da amostra e p o número de coeficientes do modelo sob-pesquisa. Outro ponto importante é o fato de que quanto maior é o número de regressores, menor será a *Deviance*, porém, o princípio da parcimônia indica que ter muitos regressores no modelo não o melhor para o seu ajuste.

Em resumo, se a *Deviance* é pequena, grande parte da variabilidade dos dados é explicada pelo ajuste do Modelo Sob Pesquisa. Se a *Deviance* for moderada, temos um modelo simples e interpretável e, se a *Deviance* for grande, o ajuste do Modelo Sob Pesquisa não explica bem a variabilidade dos dados. Existem diversos métodos para verificar se a *Deviance* é pequena, moderada ou grande. Nesse trabalho, diremos que a *Deviance* será moderada quando ela apresentar valores próximos ou menores do que valor esperado de sua distribuição assintótica, ou seja, se seu valor for menor ou próximo de $(n-p)$.

Um outro método de se avaliar o ajuste do modelo é usar gráficos de *half-normal* de um modelo de diagnóstico sendo possível utilizar de diferentes tipos de resíduos. Tais gráficos

podem ser realizados “plotando os valores absolutos ordenados de um modelo versus a estatística de ordem esperada de uma distribuição *half-normal*” (Moral e Demétrio (2017), tradução pessoal).

Adicionado a esse gráfico, temos um envelope simulado, proposto por Atkinson (1985), que, se o modelo apresentar bom ajuste, os pontos do gráfico provavelmente irão estar dentro do envelope. O objetivo desse envelope é ajudar na definição se o ajuste do modelo foi bom ou não, fazendo uso da quantidade de pontos que temos dentro/fora do envelope. O pacote `Hnp` do R tem uma função que realiza esse gráfico com envelope simulado.

3.4 Células Tarifárias

Ohlsson e Johansson (2010) afirmam que, no caso de haver um número de sinistros suficientes em cada célula tarifária, que é um subconjunto de características observáveis do veículo segurado e do condutor, seria possível estimar o prêmio para cada célula estimando-se o custo esperado pelo prêmio puro observado. Entretanto, na prática, o número de sinistros geralmente não é suficiente. Sendo assim, “há uma necessidade de métodos que forneçam um prêmio puro esperado que varia mais suavemente ao longo das células, com boa precisão das estimativas das células” (Ohlsson e Johansson (2010), tradução pessoal). Um método com as características citadas anteriormente pelos autores são os Modelos Multiplicativos.

Tal método apura como o prêmio puro depende de uma série de fatores de risco (covariáveis). Considere que temos M fatores de risco, cada um com suas m_i classes (categorias). Por simplicidade, considere o caso $M = 2$. Nesse caso, (i, j) denota a célula tarifária, onde i e j representam a classe do primeiro e segundo fator de risco, respectivamente. Considere que, para tal célula, tem-se a exposição w_{ij} que, na definição de Ohlsson e Johansson (2010), é uma medida de volume, que pode ser duração, número de sinistros, prêmio ganho, entre outros. Além disso, também tem-se a resposta Z_{ij} definida por Ohlsson e Johansson (2010) como o resultado de uma variável aleatória que pode ser o número de sinistro, o custo dos sinistros, entre outros. A resposta e a exposição da célula conduz ao índice-chave que é a proporção entre a resposta e a exposição, ou seja,

$$Y_{ij} = \frac{Z_{ij}}{w_{ij}}$$

O valor esperado do índice-chave é definido como:

$$E(Y_{ij}) = \mu_{ij}$$

mais detalhes sobre essa definição pode ser encontrado em Ohlsson e Johansson (2010).

Sendo assim, o modelo multiplicativo é definido da seguinte forma:

$$\mu_{ij} = \gamma_0 \gamma_{1i} \gamma_{2j}$$

onde γ_0 é o valor de base (índice-chave para as apólices na célula de base, também denominado de *baseline*), γ_{1i} , $i = 1, \dots, m_1$, são os parâmetros correspondentes às diferentes classes do primeiro fator de risco e γ_{2j} , $j = 1, \dots, m_2$, são os parâmetros correspondentes às diferentes classes do segundo fator de risco. γ_{1i} e γ_{2j} medem as diferenças relativas ao *baseline*. Exemplificando, considere que a célula (1, 1) corresponde ao *baseline* e $\gamma_{12} = 1,5$. Logo, a média da célula (1, 2) é 50% maior que na célula de base (1, 1).

Finalmente, o caso geral dos modelos multiplicativos pode ser formulado assim:

$$\mu_{ij} = \gamma_0 \gamma_{1i_1} \gamma_{2i_2} \dots \gamma_{Mi_M}$$

Toda a definição acima foi feita para estimar o prêmio puro, mas pode ser aplicada para frequência e severidade.

3.5 Modelo Tweedie

Na seção 3.3.1, foi exposto que a Família Exponencial engloba distribuições discretas e contínuas. Entretanto, também temos nessa família distribuições que são misturas de componentes discretos e contínuos.

Uma distribuição de mistura utilizada para precificar o Prêmio de Risco de um seguro é a distribuição Tweedie que tem uma massa de probabilidade positiva no zero, representando os casos sem indenizações, e um componente contínuo para valores positivos representando as indenizações. Tal distribuição pode ser caracterizada como uma distribuição composta de Poisson. Além disso, a Tweedie é uma soma aleatória de variáveis aleatórias Gama representando perdas individuais. A variável resultante será denominada de perda agregada (Xacur & Garrido (2015)).

Levando em consideração que muitas apólices permitem múltiplos sinistros, seja N o número de sinistros de uma apólice e Y_1, \dots, Y_N o valor individual dos sinistros. Sendo assim, o valor total dos sinistros é: $S = Y_1 + \dots + Y_N$, onde $S = 0$ se $N = 0$. Portanto, se N tem distribuição de Poisson(λ) e Y_i 's são variáveis aleatórias independentes com distribuição Gama($\alpha; \theta$), temos que S tem distribuição Tweedie.

Essa distribuição será utilizada para fazer o ajuste de um modelo unificado para frequência e severidade, e comparar com o ajuste feito separadamente.

4 Resultados

4.1 Análise Descritiva

Na Figura 4.1, temos um conjunto de gráficos que mostra a distribuição da exposição e quantidade de sinistros (de acordo com o tipo de cobertura) por região.

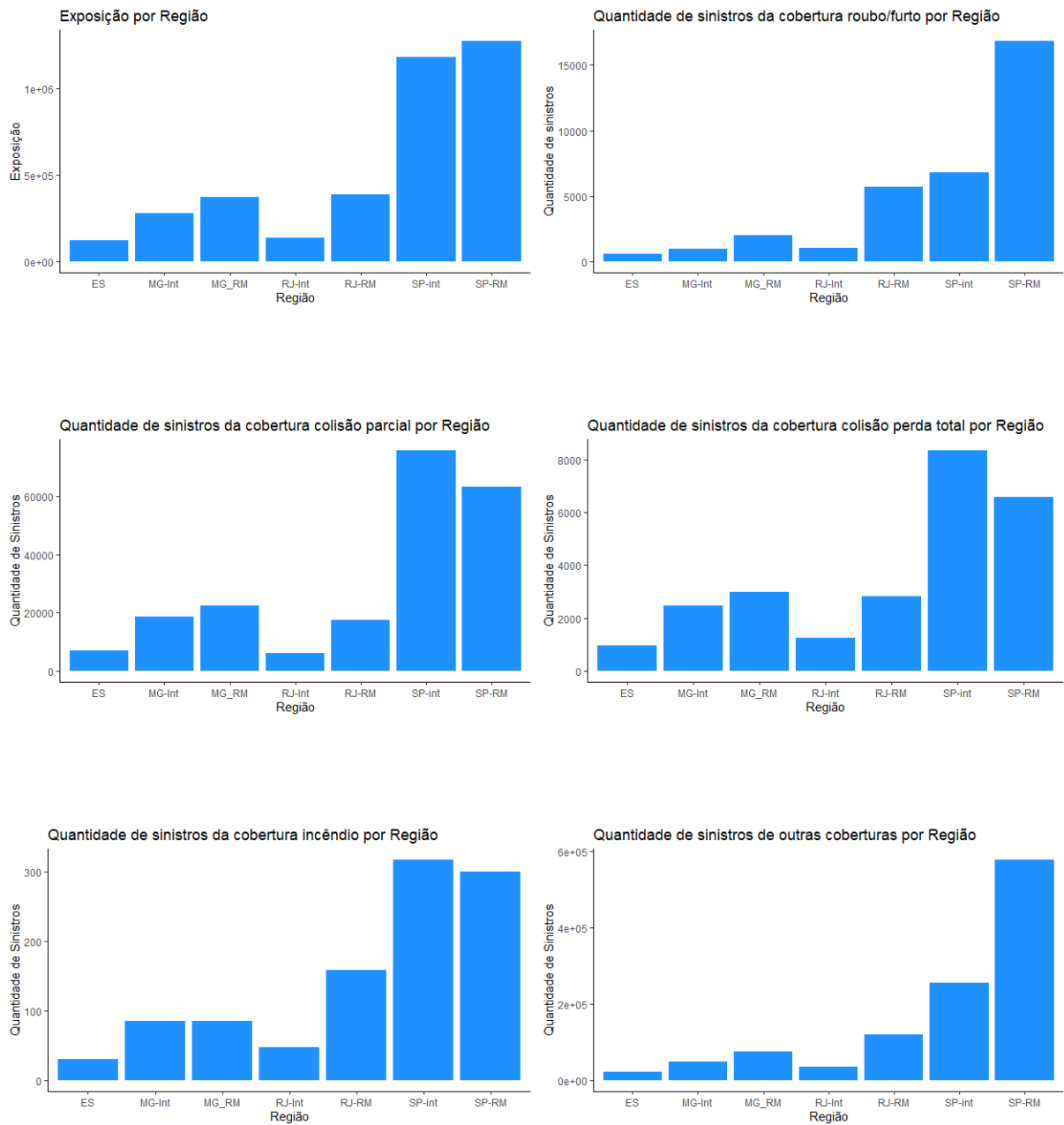


Figura 4.1 – Gráficos de exposição e quantidade de sinistros (de acordo com o tipo de cobertura) por região

Analisando os gráficos da Figura 4.1, pode-se notar que a exposição no segundo semestre de 2019 foi muito maior em São Paulo do que nos outros Estados (tanto para Região Metropolitana quanto para Interior). Além disso, Rio de Janeiro - Região Metropolitana e Minas Gerais - Região Metropolitana apresentam patamares similares quanto à exposição semestral, seguidos por Minas Gerais - Interior, Rio de Janeiro - Interior e Espírito Santo.

Já nos gráficos de quantidade de sinistros no segundo semestre de 2019 por tipo de cobertura, podemos destacar o fato de que, com exceção aos sinistros da cobertura roubo/furto e outras coberturas, São Paulo - Interior apresentou quantidade de sinistros maior do que São Paulo - Região Metropolitana e, isso pode ser explicado pelo fato de que o Interior está englobando uma área maior do que a Região Metropolitana.

Tal discrepância observada entre São Paulo e os outros Estados quanto à exposição semestral e quantidade de sinistros (em valores absolutos) já era esperada, visto que 59% da frota de veículos da Região Sudeste se encontra em São Paulo⁷. Conseqüentemente, esperamos que o valor do prêmio em São Paulo (Interior e Região Metropolitana) seja maior do que nas outras regiões.

Tabela 4.1 – Regiões e suas respectivas frequências de sinistros por tipo de cobertura

Região	Exposição	Roubo/ Furto	Colisão Parcial	Colisão Perda Total	Incêndios	Outras Cober- turas
ES	123.244	0,00465	0,0572	0,00780	0,000243	0,180
MG - Int	278.603	0,00338	0,0672	0,00891	0,000305	0,178
MG- RM	371.460	0,00532	0,0606	0,00807	0,000229	0,202
RJ - Int	137.103	0,00748	0,0447	0,00902	0,000350	0,263
RJ - RM	385.126	0,01480	0,0457	0,00732	0,000413	0,310
SP - Int	1.182.700	0,00577	0,0641	0,00705	0,000268	0,215
SP - RM	1.274.603	0,01320	0,0495	0,00516	0,000235	0,453

Utilizando a definição de frequência que se encontra no Capítulo 3 (quantidade de sinistros dividido pela exposição), a Tabela 4.1 mostra que a frequência no segundo semestre de 2019 para sinistros das coberturas roubo/furto e incêndio é maior na Região Metropolitana do RJ. Para cobertura colisão parcial, a maior frequência semestral está na região Interior de MG. Já para colisão perda total, nota-se que a maior frequência semestral se encontra em RJ - Interior e, por fim, para outras coberturas, a maior está em SP - Região Metropolitana.

Nas seções seguintes, analisaremos separadamente a frequência e severidade dos sinistros para cada tipo de cobertura.

⁷ Ministério da Infraestrutura. **Frota de Veículos - 2021**. Disponível em <https://www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-denatran/frota-de-veiculos-2021>
Acesso em: 20 de junho de 2021

4.1.1 Sinistros da cobertura roubo/furto

Vamos analisar as frequências de sinistros apenas relacionados a sinistros de roubo e furto, para os diferentes códigos tarifários e ano do modelo dos veículos. Pelo gráfico da Figura 4.2, podemos ver que a frequência de roubo/furto de motocicletas é aproximadamente três vezes maior do que a frequência semestral para os outros tipos de códigos tarifários, sendo que estes estão no mesmo patamar. Já a Figura 4.3 indica que não há diferença significativa entre os anos de modelo com respeito à frequência semestral de roubo/furto.

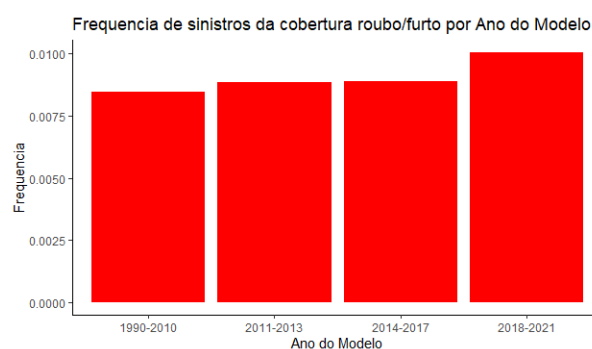
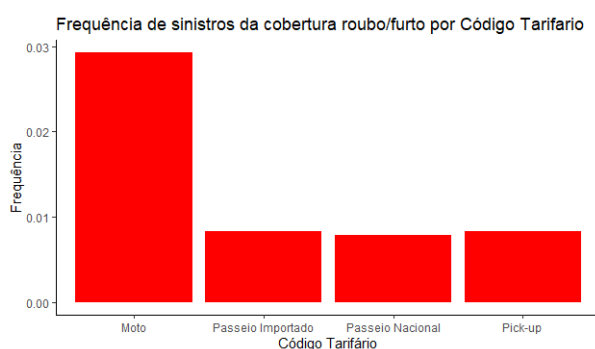


Figura 4.2 – Frequência 1 por cód. tarif. Figura 4.3 – Frequência 1 por ano modelo

Analisando quanto à idade do motorista, através do gráfico da Figura 4.4, notamos que a frequência semestral de roubo/furto apresenta queda nas faixas de idade mais velhas com relação às faixas mais novas. E na Figura 4.5 vemos que a frequência é maior para o sexo masculino em todas as regiões, e que SP - Região Metropolitana e RJ - Região Metropolitana apresentam frequência semestral de roubo/furto maior do que nas outras regiões.

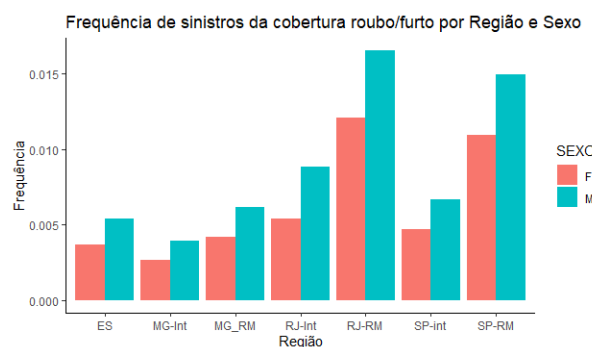
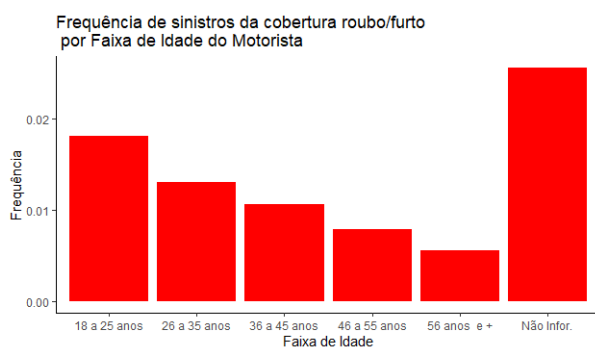


Figura 4.4 – Frequência 1 por idade Figura 4.5 – Frequência 1 por região e sexo

Quanto à severidade, definida como a divisão entre o valor da indenização e a quantidade de sinistros, vemos pela Figura 4.6 que a severidade no segundo semestre de 2019 dos sinistros da cobertura roubo/furto foi maior para a categoria Pick-up, seguida por Passeio Importado, Passeio Nacional e Motocicletas. Com relação ao ano do modelo, explicitado

no gráfico da Figura 4.7, nota-se que quanto mais novo o carro, maior foi a severidade semestral, com veículos das faixas 2014 a 2017 e 2018 a 2021 apresentando valores muito próximos.

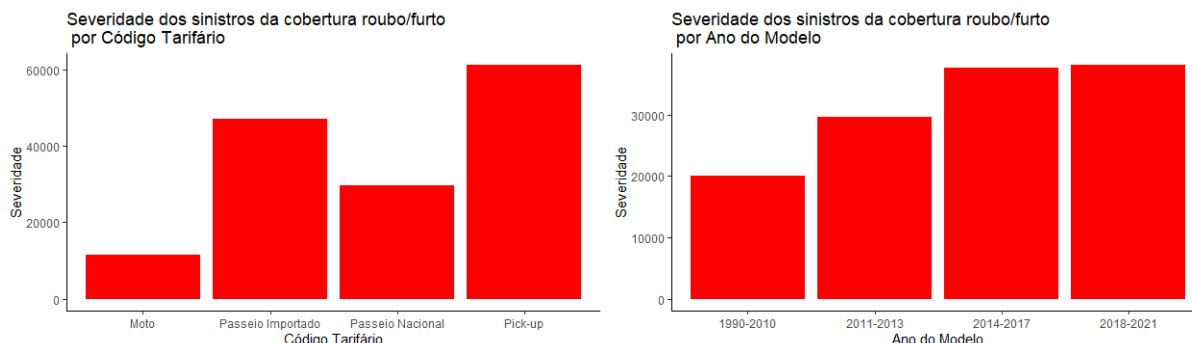


Figura 4.6 – Severidade 1 por cód. tarif. Figura 4.7 – Severidade 1 por ano modelo

Analisando a severidade em relação à idade do motorista, apresentado no gráfico da Figura 4.8, podemos ver que quanto mais velho é o condutor do veículo, maior foi a severidade semestral de furto/roubo. Além disso, nota-se uma severidade alta para o grupo Idade Não Informada. E, pelo gráfico da Figura 4.9, as duas regiões de São Paulo, seguidas por MG - Interior, apresentaram as menores severidades semestral de sinistros da cobertura roubo/furto. As demais regiões apresentaram valores próximos no que diz respeito à severidade.

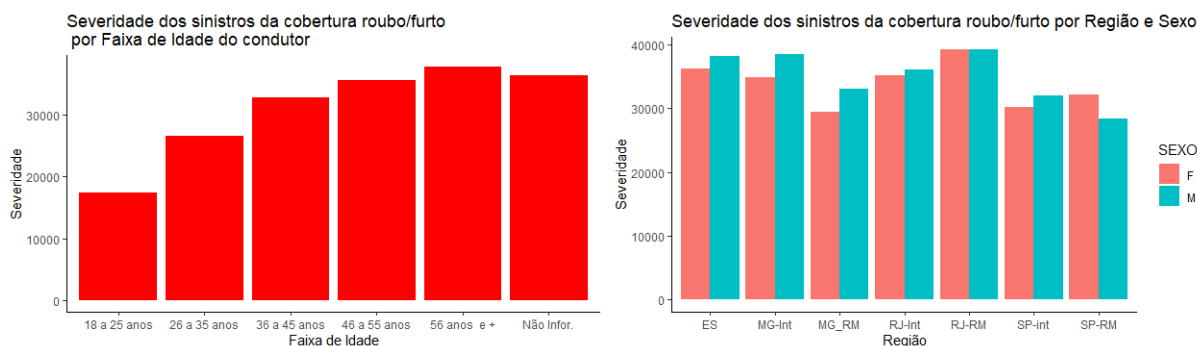


Figura 4.8 – Severidade 1 por idade Figura 4.9 – Severidade 1 por região e sexo

Apesar de existir muitas diferenças entre as regiões no que tange à frequência semestral de sinistros da cobertura roubo/furto, como notado na Figura 4.5, a severidade apresenta pequenas variações entre as localidades, indicando que, não importa onde o veículo foi roubado/furtado, que o valor da perda vai ser semelhante, com RJ - Região Metropolitana sendo levemente superior às demais. Além disso, em SP - Região Metropolitana, a severidade em estudo foi maior para o sexo feminino do que a do sexo masculino e, nas outras regiões, o oposto aconteceu.

4.1.2 Sinistros da cobertura colisão parcial

Agora, veremos algumas análises à respeito de sinistros da cobertura colisão parcial. Como pode ser visto através do gráfico da Figura 4.10, a frequência semestral de sinistros da cobertura parcial é maior na categoria Pick-up, seguido de Passeio Nacional e Passeio Importado, e com as Motocicletas apresentando o menor índice. Quanto ao Ano do Modelo, presente no gráfico da Figura 4.11, a frequência de ocorrência de colisão parcial no segundo semestre de 2019 apresentou valores crescentes quando se olha dos veículos mais antigos para os mais novos.

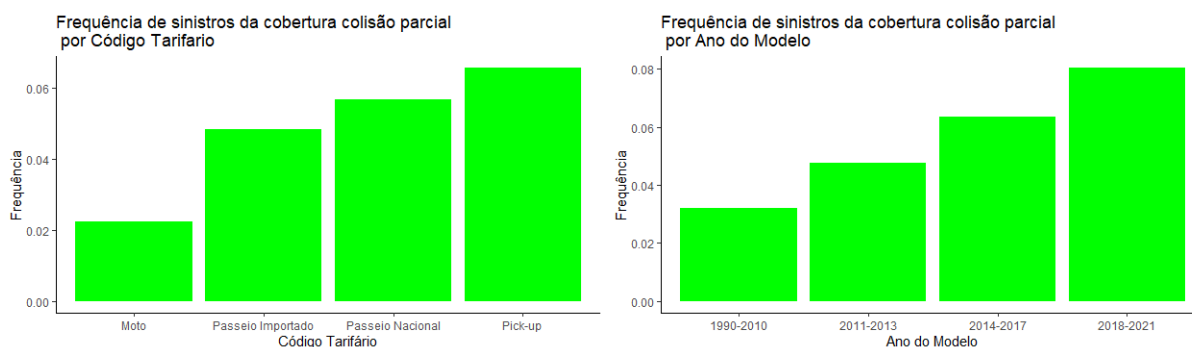


Figura 4.10 – Frequência 2 por cód. tarif. Figura 4.11 – Frequência 2 por ano modelo

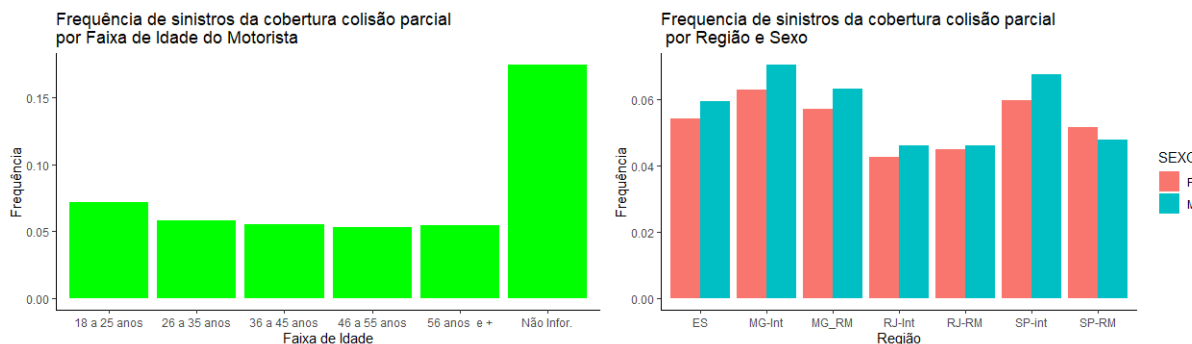


Figura 4.12 – Frequência 2 por idade Figura 4.13 – Frequência 2 por região e sexo

Como demonstrado pelo gráfico da Figura 4.12, à exceção da Idade Não Informada, não temos muita diferença entre as frequências de colisão parcial entre as faixas etárias, sendo de 18 a 25 anos levemente maior que as outras. E, quanto à região e sexo, podemos notar na Figura 4.13 que, as duas regiões de MG e SP - Região Metropolitana apresentaram o maior número de ocorrências do sinistro em estudo e, com exceção de São Paulo - Região Metropolitana, a frequência no semestre em estudo foi sempre maior para o sexo masculino.

Avaliando agora a severidade semestral dos sinistros da cobertura colisão parcial, nota-se que, quanto ao código tarifário do veículo, exibido na Figura 4.14, a maior severidade se encontra na categoria Passeio Importado seguida por Pick-up. Ademais, Motocicleta

e Passeio Nacional apresentam valores semelhantes. No que tange ao ano do modelo, o gráfico da Figura 4.15 mostra que a severidade semestral de sinistros da cobertura colisão parcial é crescente à medida que o ano do veículo fica mais novo.

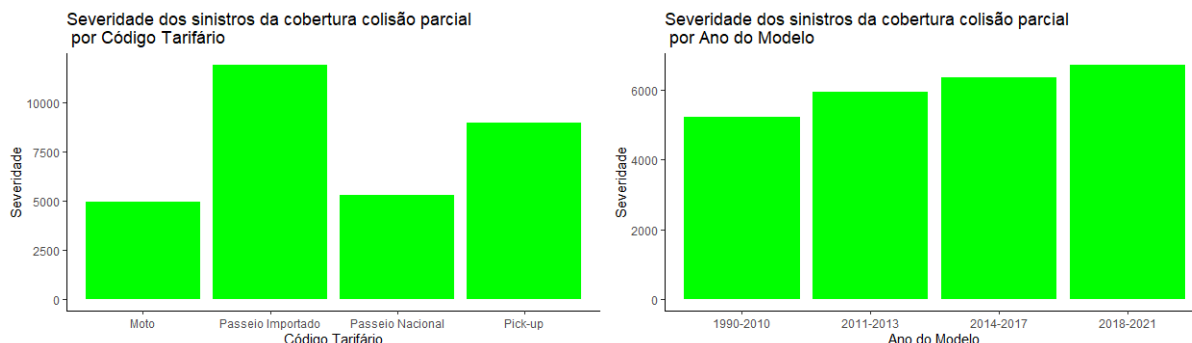


Figura 4.14 – Severidade 2 por cód. tarif. Figura 4.15 – Severidade 2 por ano modelo

Em relação à idade do condutor, nota-se pelo gráfico da Figura 4.16 que temos pouca diferença no valor da perda entre as faixas de idade do condutor, com a faixa de 36 a 45 anos apresentando valor levemente superior às demais faixas. Entretanto, a severidade no semestre em estudo é maior no grupo com Idade Não Informada. Quanto à localidade, temos pouca diferença na severidade semestral entre as regiões, com SP- Interior apresentando menor valor, como pode ser visto na Figura 4.17. Além disso, o gráfico da figura citada anteriormente mostra que, sem exceções, a severidade no semestre em estudo é maior para o sexo masculino.

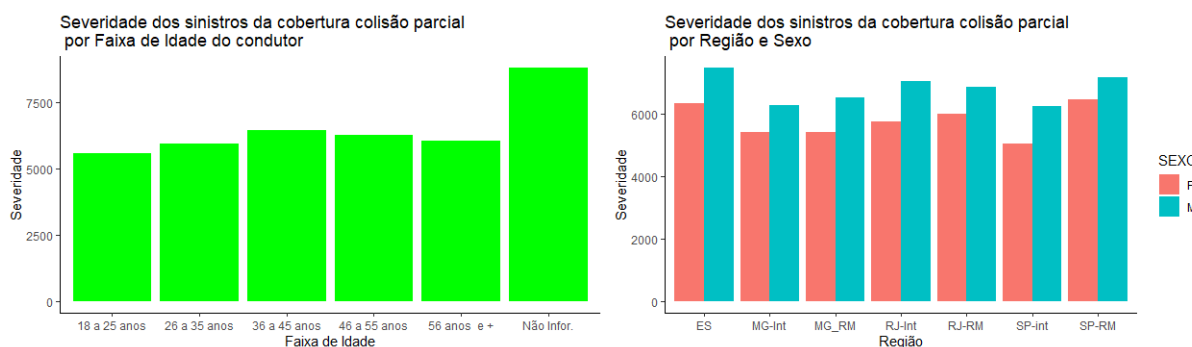


Figura 4.16 – Severidade 2 por idade Figura 4.17 – Severidade 2 por região e sexo

4.1.3 Sinistros da cobertura colisão perda total

Através da Figura 4.18, nota-se que os veículos do tipo Passeio Importado e Moto apresentaram frequência semestral de sinistros de colisão perda total maiores do que as outras duas categorias. Ademais, a Figura 4.19 indica que, quanto mais novo é o modelo do carro, menor a frequência semestral de sinistros da cobertura perda total.

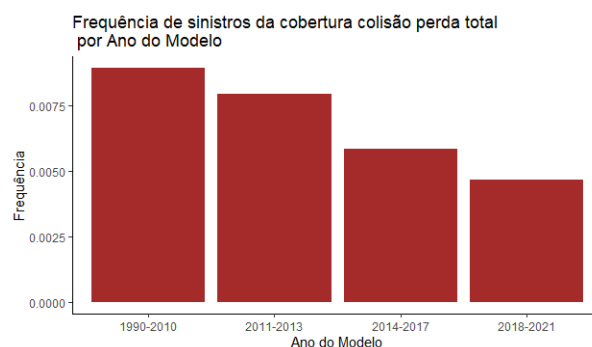
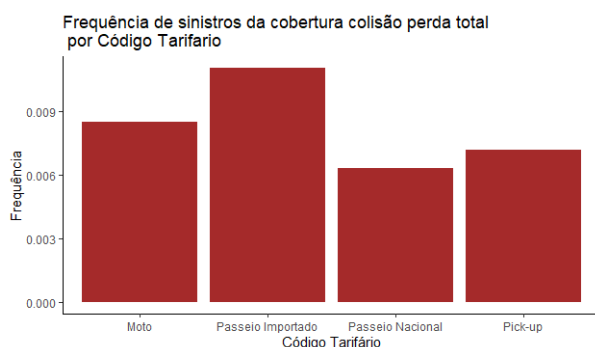


Figura 4.18 – Frequência 3 por cód. tarif. Figura 4.19 – Frequência 3 por ano modelo

Na Figura 4.20, pode-se perceber que a maior frequência semestral de colisão perda total é para Idade Não Informada, seguida pela faixa de 18 a 25 anos. Além disso, as demais faixas apresentaram valores semelhantes. E, para região e sexo, a Figura 4.21 mostra a prevalência da frequência em estudo para o sexo masculino em todas as regiões e que, essa frequência apresenta, à exceção de São Paulo - Região Metropolitana, valores similares quanto ao sexo e à localidade.

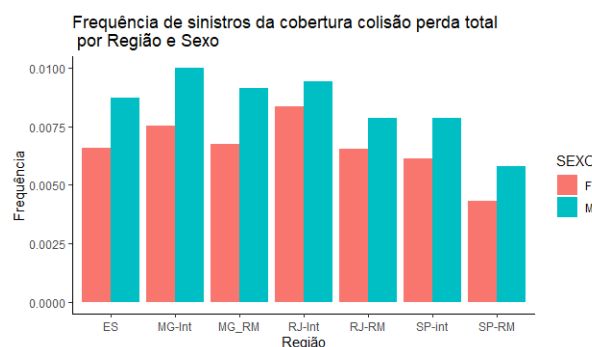
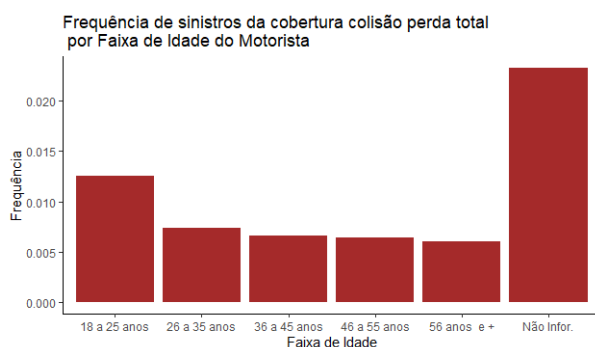


Figura 4.20 – Frequência 3 por idade Figura 4.21 – Frequência 3 por região e sexo

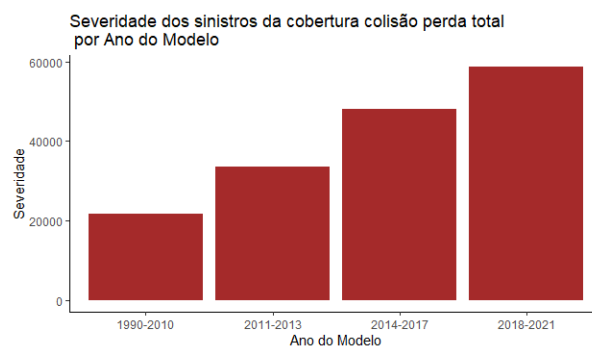
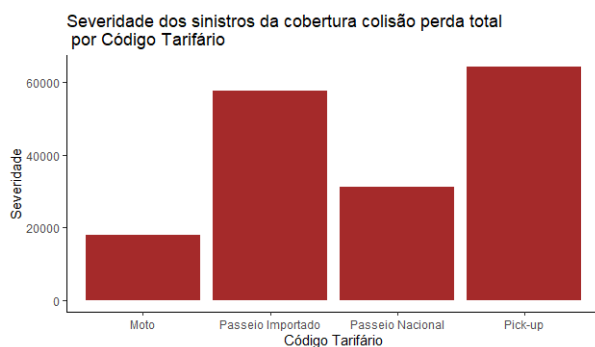


Figura 4.22 – Severidade 3 por cód. tarif. Figura 4.23 – Severidade 3 por ano modelo

Os sinistros da cobertura perda total apresentaram severidade maior para veículos do tipo Pick-up, seguido por Passeio Importado, e o menor valor é apresentado pela categoria

Motocicleta. Esses fatos podem ser identificados no gráfico da Figura 4.22. Já na Figura 4.23 nota-se que, quanto mais novo é o veículo, maior é a sua severidade no semestre em estudo.

No que diz respeito à idade do motorista, o gráfico da Figura 4.24 mostra que as duas primeiras faixas etárias apresentaram valores mais baixos para a severidade semestral de sinistros da cobertura perda total, se comparado com as outras faixas que, incluindo Idade Não Informada, apresentaram valores próximos para a perda média. Quanto à região, Espírito Santo apresentou maior severidade no semestre em estudo, seguido por MG - Interior. As demais regiões apresentaram severidade em patamares similares. No que diz respeito ao sexo, com exceção de SP - Região Metropolitana, em todas as regiões a severidade foi maior para o sexo masculino.

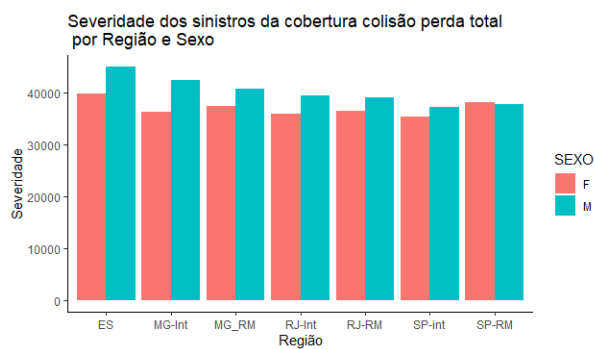
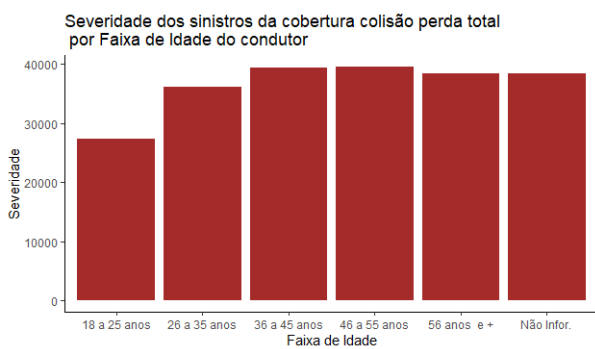


Figura 4.24 – Severidade 3 por idade Figura 4.25 – Severidade 3 por região e sexo

4.1.4 Sinistros da cobertura incêndios

Agora vamos analisar os sinistros da cobertura incêndios. Para esse tipo de cobertura, as Pick-ups apresentaram frequência no semestre em estudo maior do que Passeio Nacional e Passeio Importado, como demonstrado na Figura 4.26, com Motocicleta apresentando uma frequência irrisória perto as outras categorias. Ademais, através da Figura 4.27 pode ser visto que quanto mais novo o modelo do veículo, menor é a frequência semestral de sinistros da cobertura incêndios.

Já para a idade do condutor, apresentada na Figura 4.28, vemos que a frequência semestral de incêndios tem valor similar entre as faixas etárias, entretanto, para idade não informada, a frequência foi muito maior do que para as demais faixas. Quanto à região e sexo, a Figura 4.29 demonstra que Espírito Santo, MG - Região Metropolitana e SP - Região Metropolitana apresentaram patamares próximos da frequência em estudo e essa frequência é similar para os dois sexos. Já para as demais regiões, temos valores maiores para a frequência semestral de incêndio com discrepâncias grandes entre os sexos masculino e feminino.

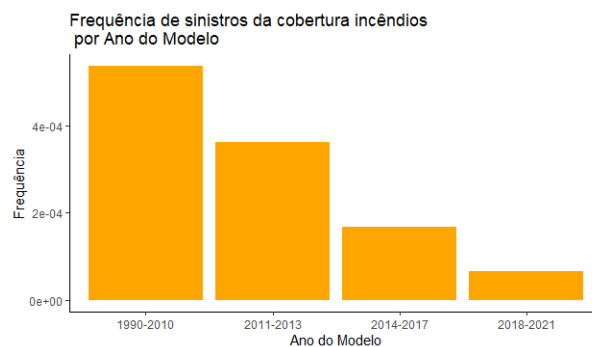
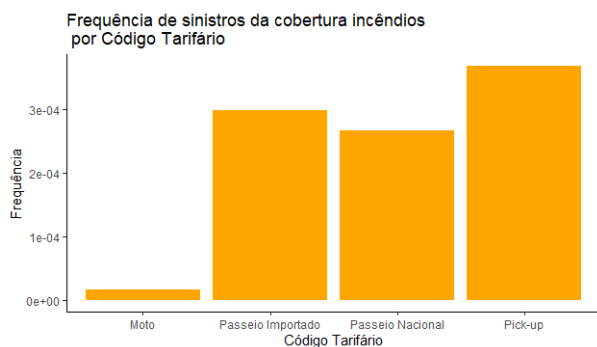


Figura 4.26 – Frequência 4 por cód. tarif. Figura 4.27 – Frequência 4 por ano modelo

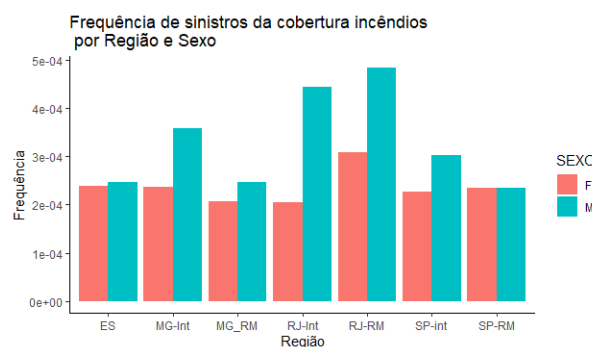
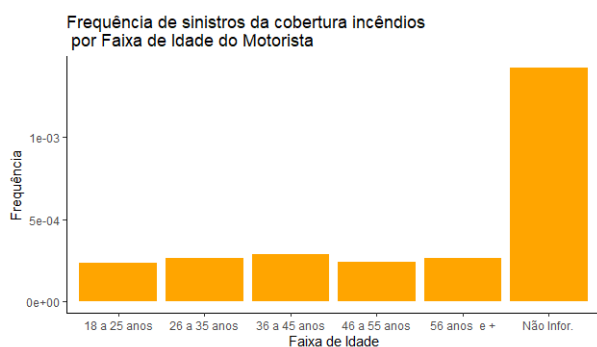


Figura 4.28 – Frequência 4 por idade Figura 4.29 – Frequência 4 por região e sexo

No que diz à respeito da severidade no semestre em estudo, os veículos do tipo Pick-up apresentaram o maior total de indenizações por sinistros da cobertura incêndios, seguido pelo Passeio Importado, como explicitado no gráfico da Figura 4.30. Passeio Nacional e Motocicletas apresentaram valores baixos para a severidade em estudo. Assim como observado para a frequência, podemos ver que a severidade aumenta à medida que a idade do veículo diminui, como podemos ver na Figura 4.31.

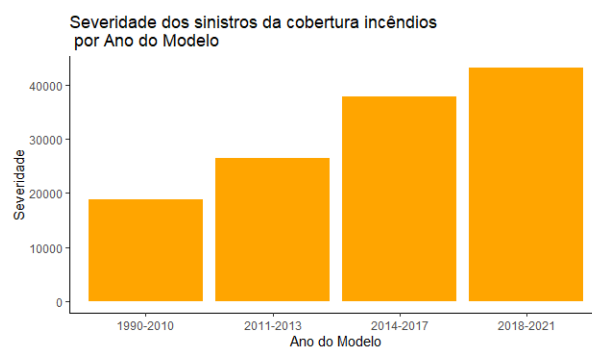
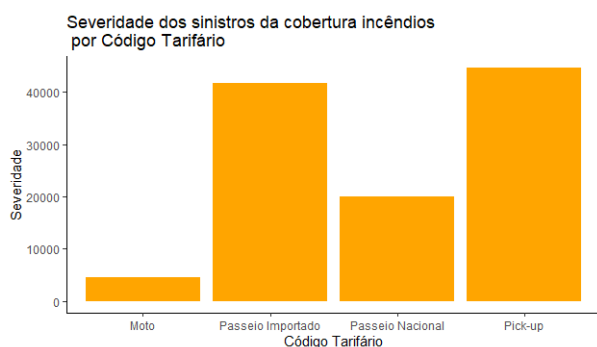


Figura 4.30 – Severidade 4 por cód. tarif. Figura 4.31 – Severidade 4 por ano modelo

No que diz respeito à idade do condutor apresentado na Figura 4.32, a severidade de sinistros da cobertura incêndios apresentou pouca diferença entre os grupos de idade do motorista. Para região e sexo, vemos na Figura 4.33 que a única localização na qual a

severidade em estudo para o sexo feminino foi maior do que para o sexo masculino foi em MG - Interior, sendo que, na frequência (apresentado na Figura 4.29) a incidência foi maior para o sexo masculino nessa região. Para as demais regiões, a severidade semestral é maior para o sexo masculino. Ademais, a perda média é maior para MG e ES do que para as demais regiões.

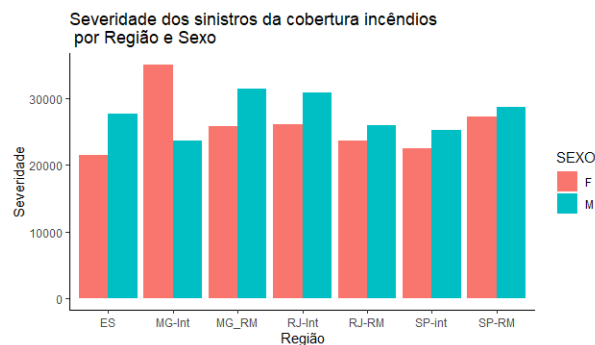
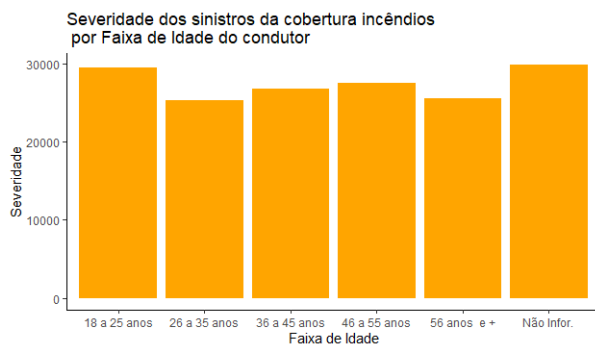


Figura 4.32 – Severidade 4 por idade

Figura 4.33 – Severidade 4 por região e sexo

4.1.5 Sinistros de outras coberturas

E, por fim, para sinistros de outras coberturas, que engloba assistência 24 horas e outros serviços, vemos através da Figura 4.34 que, com exceção de Moto, que apresentou frequência baixa no semestre em estudo, os outros códigos tarifários apresentaram frequências similares. No que diz respeito ao ano do modelo, novamente vemos que, quanto mais novo o carro, menor é a frequência semestral de sinistros de outras categorias, como explicitado pela Figura 4.35.

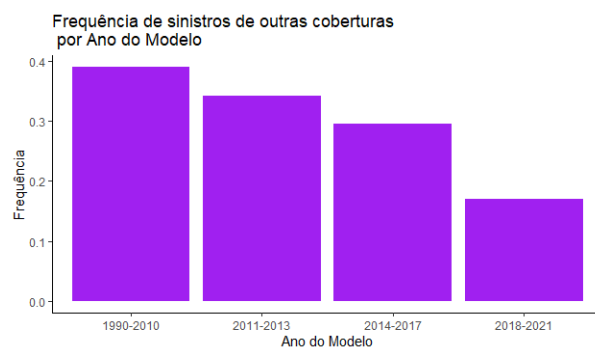
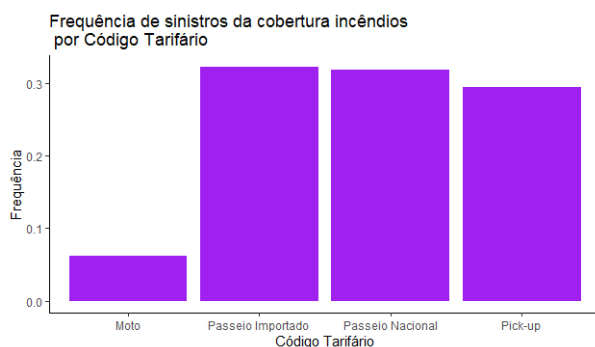


Figura 4.34 – Frequência 9 por cód. tarif.

Figura 4.35 – Frequência 9 por ano modelo

No que diz respeito à faixa de idade de motorista, mostrado pelo gráfico da Figura 4.36, vemos que as faixas Idade Não Informada e 18 a 25 anos apresentaram as menores frequências de sinistros de outras coberturas no semestre em estudo, e as demais faixas tiveram frequências no mesmo patamar entre elas.

Já para região e sexo, explicitados no gráfico da Figura 4.37, temos que, à exceção de Espírito Santo, em todas as outras regiões a frequência semestral de sinistros de outras coberturas foi maior para o sexo feminino. Ademais, São Paulo - Região Metropolitana e Rio de Janeiro - Região Metropolitana apresentaram as maiores frequências, com as outras regiões tendo valores no mesmo patamar.

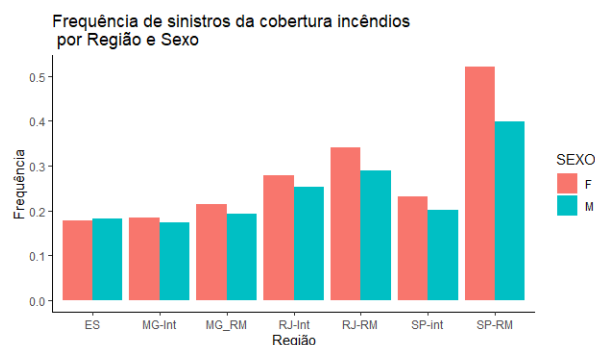
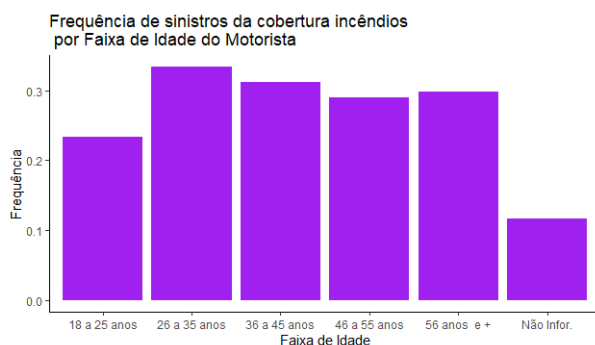


Figura 4.36 – Frequência 9 por idade Figura 4.37 – Frequência 9 por região e sexo

Quanto à severidade no semestre em estudo para sinistros de outras coberturas, o gráfico da Figura 4.38 mostra que o comportamento da perda média em relação ao código tarifário é semelhante ao comportamento apresentado para a severidade semestral de sinistros da cobertura colisão parcial (Figura 4.14). Além disso, como pode ser notado no gráfico da Figura 4.39, a severidade é maior para veículos com ano do modelo na faixa de 2018 a 2021, com modelos dos anos 2011 a 2013 e 2014 a 2017 apresentando números próximos.

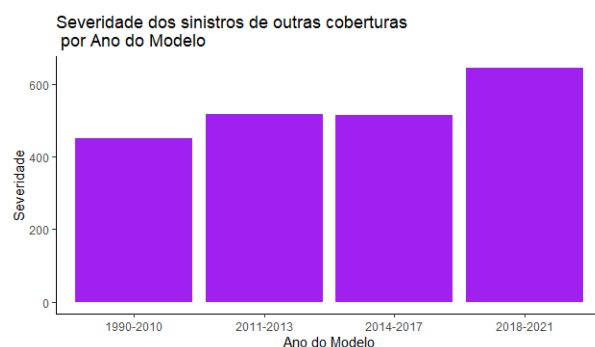
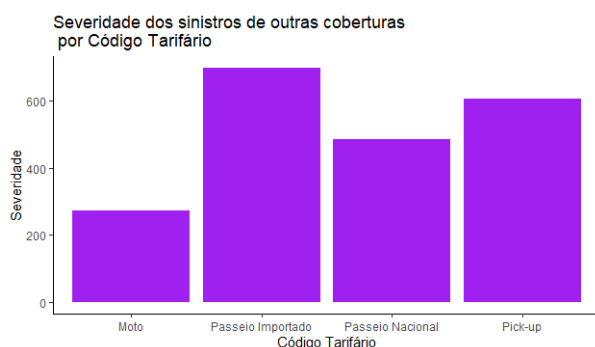


Figura 4.38 – Severidade 9 por cód. tarif. Figura 4.39 – Severidade 9 por ano modelo

Além disso, a Figura 4.40 destaca a imensa prevalência da severidade semestral no grupo Idade Não Informada, com as outras faixas apresentando valores irrisórios perto desse grupo. Por fim, apesar da frequência do sinistro em estudo ter sido maior para o sexo feminino, a severidade semestral, apresentada na Figura 4.41, é maior para o sexo masculino. Ainda dentro dessa comparação com a frequência (Figura 4.37), vemos que, quanto menor foi a frequência de sinistros de outra cobertura em uma região, maior foi sua severidade.

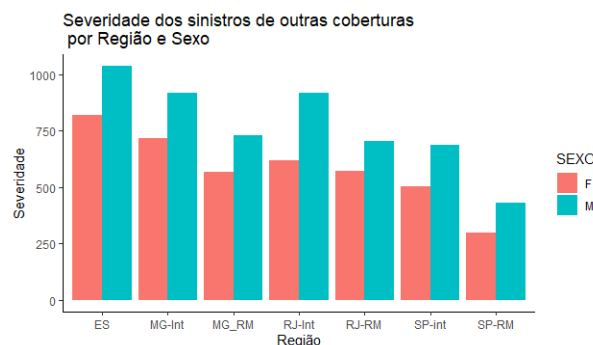
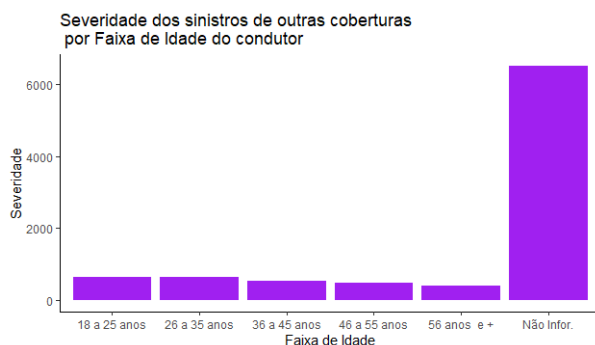


Figura 4.40 – Severidade 9 por idade Figura 4.41 – Severidade 9 por região e sexo

4.2 Ajuste e interpretação dos modelos para Frequência e Severidade

O ponto de partida para modelagem foi usar o banco de dados final completo. Sendo assim, para a severidade, foram testadas as distribuições Gama e Log-Normal. Vale ressaltar que, quando modela-se a severidade, são retirados todos as observações cuja severidade seja zero. Ademais, o *log* da exposição é incluído no ajuste como um *offset*, ou seja, a exposição será uma variável explicativa, mas não será estimado um coeficiente para ela. Isso posto, para todos os tipos de cobertura, a modelagem usando ambas as distribuições apresentaram todas as covariáveis significativas e gráficos do pacote `Hnp` com quase todos os pontos fora do intervalo. Entretanto, os valores da *Deviance* do modelo Gama são menores do que os graus de liberdade da distribuição Qui-Quadrado, enquanto os valores da *Deviance* do modelo Log-Normal estavam muito próximos dos graus de liberdade. Apesar dessa diferença, pode-se escolher trabalhar com ambas as distribuições devido ao critério definido na subseção 3.3.4. Vale destacar que, como estamos falando de distribuições que possuem parâmetro de dispersão (Poisson, Gama, Log-Normal, Tweedie) a *Deviance* será tratada como *Deviance* escalonada. Portanto quando *Deviance* é mencionada, estamos tratando da medida escalonada.

Já para frequência, foram testados modelos usando três distribuições: Poisson, Binomial Negativa e Log-Normal. É importante destacar que, devido à média da quantidade de sinistros para todos os tipos de cobertura ser maior do que 5, é possível utilizar a distribuição contínua Log-Normal para dados contagem (esse resultado é apresentado por Montgomery e Runger (2016)). Para todos os tipos de coberturas, as modelagens usando Poisson e Binomial Negativa apresentaram quase nenhuma ou nenhuma covariável significativa. Ademais, os valores de suas *Deviances* eram muito baixos e apresentaram gráficos do pacote `Hnp` com quase todos os pontos fora do intervalo. Portanto, calcular o prêmio usando essas distribuições não seria uma escolha interessante.

No modelamento usando a distribuição Log-Normal, todas as variáveis explicativas foram significativas, acompanhado pela *Deviance* apresentando valores muito próximos

dos graus de liberdade da distribuição Qui-quadrado. Entretanto, o comportamento dos gráficos do pacote **Hnp** foi similar ao comportamento das outras distribuições. Apesar disso, a escolha pelo modelo usando a Log-Normal é interessante.

Entretanto, o resultado encontrado para as células tarifárias usando o modelo multiplicativo é muito próximo de zero. Isso pode ser explicado pelo fato de ter muita frequência igual a zero. A título de exemplo, para cobertura roubo/furto, até o primeiro quartil encontramos frequência zero, isso quer dizer que, no mínimo 25% dos dados de roubo/furto apresentam frequência zerada. Ou seja, estamos modelando muitos “zeros”. E, mesmo após multiplicar os resultados das células tarifárias para frequência pelos resultados das células tarifárias para severidade, encontra-se muitos prêmios iguais a zero, valor esse que não é interessante para uma seguradora.

Outro ponto que vale ser destacado é a categoria Idade Não Informada. Não temos informações na base de dados a cerca dessa categoria de idade, além de que quando da contratação do seguro é informado a idade do motorista. Portanto, faz-se necessário a retirada dessa categoria na hora de modelar.

Sendo assim, optou-se por realizar novos modelos retirando do banco de dados observações com frequência igual a zero e observações cuja Idade é Não Informada.

Mesmo com essa abordagem, a modelagem para frequência usando as distribuições Poisson e Binomial Negativa apresentou os mesmos problemas citados anteriormente (quase nenhuma ou nenhuma covariável significativa, *Deviance* muito baixa, **Hnp** com quase todos os pontos fora do envelope). Novamente, a escolha por um modelo usando uma dessas duas distribuições não é interessante.

Agora, usando a distribuição Log-Normal, nos modelos para roubo/furto, colisão perda total e outras coberturas, todas as covariáveis foram significativas, *Deviances* próximas dos graus de liberdade da Qui-Quadrado, e gráficos do pacote **Hnp** com entre 60% a 99% dos pontos fora do intervalo. No modelo para cobertura incêndios e no modelo para cobertura colisão parcial, a covariável sexo não foi significativa. Retirando sexo do modelo, encontra-se todas as covariáveis significativas, *Deviance* próxima dos graus de liberdade da Qui-Quadrado e todos os pontos dentro do envelope, como apresentado na Figura 4.42.

Já na modelagem da severidade, visto que a retirada de frequência igual a zero não faz diferença nesse modelo, já que se a frequência é zero, a severidade também o é, só foi feita a retirada da categoria Idade Não Informada. Apesar dessa exclusão, o comportamento do ajuste do modelo (via distribuição Gama ou Log-Normal) não diferiu muito do que foi ajustado com a categoria Idade Não Informada. Ou seja, todas as covariáveis significativas, *Deviance* menor que os graus de liberdade da distribuição Qui-Quadrado (para o modelo Gama) e próxima desses graus de liberdade (para o modelo Log-Normal). Entretanto, a *Deviance* do modelo para a severidade da cobertura incêndios se apresentou um pouco maior que os graus de liberdade do modelo. Em sua maioria, os gráficos do

pacote `Hnp` apresentaram quase todos os pontos fora do envelope. O melhor foi o gráfico para severidade dos sinistros da cobertura incêndios usando a distribuição Gama, que apresentou aproximadamente 59% dos pontos fora do envelope, como apresentado na Figura 4.43. Por ser mais fácil de interpretar, a escolha será por utilizar o modelo Gama para as severidades em estudo.

Para fins de interpretação, a seguir serão apresentados os coeficientes e seus respectivos valores-p, o gráfico do pacote `Hnp` e as relatividades do modelo de frequência semestral para cobertura incêndios. Em seguida, o mesmo será apresentado para o modelo de severidade semestral desta cobertura. Demais modelos estão apresentados nos Anexos A.

Tabela 4.2 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Incêndios

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5,3480	0,1989	-26,89	0,0000
COD_TARIF2-Passeio Importado	1,6767	0,1222	13,72	0,0000
COD_TARIF3-Pick-up	0,9494	0,0782	12,13	0,0000
COD_TARIF4-Moto	2,2258	0,3652	6,09	0,0000
REGIAOMG-Int	-0,4089	0,1666	-2,45	0,0147
REGIAOMG_RM	-0,6882	0,1645	-4,18	0,0000
REGIAORJ-Int	0,3146	0,1792	1,76	0,0801
REGIAORJ_RM	-0,3101	0,1583	-1,96	0,0511
REGIAOSP-int	-1,0286	0,1546	-6,65	0,0000
REGIAOSP_RM	-1,2918	0,1539	-8,39	0,0000
IDADE26 a 35 anos	-1,3028	0,1691	-7,70	0,0000
IDADE36 a 45 anos	-1,5447	0,1635	-9,45	0,0000
IDADE46 a 55 anos	-1,7099	0,1670	-10,24	0,0000
IDADE56 anos e +	-1,8793	0,1635	-11,49	0,0000
ANO_MODELO2011-2013	-0,4241	0,0919	-4,61	0,0000
ANO_MODELO2014-2017	-1,1527	0,0922	-12,50	0,0000
ANO_MODELO2018-2021	-1,2352	0,1219	-10,13	0,0000

Como pode ser visto na Tabela 4.2, com exceção da região RJ- Interior, todas as categorias das variáveis do modelo apresentaram valor-p abaixo de 0,05. Vale ressaltar que o coeficiente da variável `sexo` é zero pois foi retirada do modelo devido à sua não significância. Além disso, esse modelo apresentou 294 graus de liberdade e o valor de 291,78 para a *Deviance*. A interpretação desse modelo pode ser realizada usando a variável severidade transformada (*log*), ou voltando a variável para a escala original, ou seja, fazendo exponencial do coeficiente. Nesse trabalho escolheu-se fazer via segunda opção.

Os resultados da Tabela 4.3 são interpretados da maneira explicada na Seção 3.4. A título de exemplo, a relatividade da classe Ano Modelo 2011-2013 é de 0,6544, ou seja, a média dessa classe é aproximadamente 25% menor do que a classe *baseline* que é ano modelo 1990-2010. Ademais, a média da classe Ano Modelo 2014-2017 é aproximadamente 68% menor que a média do *baseline* e a média da classe Ano Modelo 2018-2021 é cerca

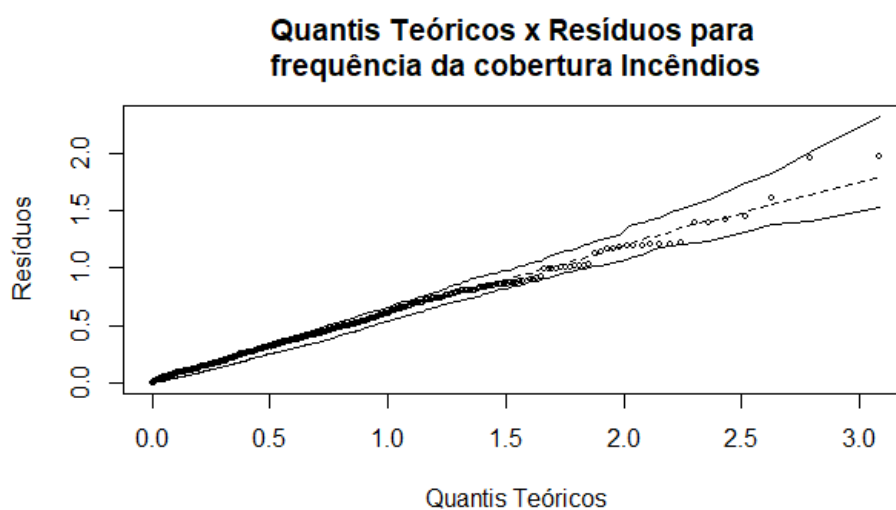


Figura 4.42 – Hnp para Frequência da cobertura Incêndios

Tabela 4.3 – Relatividades para Frequência da cobertura Incêndios

Fator de Risco	Classe	Frequência 4 Relatividade
Região	ES	1,0000
	MG - Int	0,6644
	MG - RM	0,5025
	RJ - Int	1,3697
	RJ - RM	0,7334
	SP - Int	0,3575
	SP - RM	0,2748
Sexo	F	1,0000
	M	1,0000
Idade	18-25	1,0000
	26-35	0,2718
	36-45	0,2134
	46-55	0,1809
	55+	0,1527
Código Tarifário	Passeio Nacional	1,0000
	Passeio Importado	5,3481
	Pick-up	2,5842
	Moto	9,2606
Ano Modelo	1990-2010	1,0000
	2011-2013	0,6544
	2014-2017	0,3158
	2018-2021	0,2908

de 71% menor do que a média do *baseline*. Esse resultado condiz com o gráfico da Figura 4.27, que mostra que quanto mais novo é o veículo, menor é a frequência de sinistros da cobertura incêndios.

No que se refere à severidade semestral dos sinistros da cobertura incêndios, a Tabela 4.4 mostra que as categorias RJ - Interior, Ano Modelo 2011-2013 e Ano Modelo 2014-2017 apresentaram valor-p acima de 0,05, ou seja, não são significativas. As demais

categorias são significativas. Esse modelo apresentou 290 graus de liberdade e *Deviance* aproximadamente igual a 356, que é um pouco maior do que os graus de liberdade do modelo.

Tabela 4.4 – Resultado do ajuste do modelo Gama para Severidade da cobertura Incêndios

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4,7143	0,2477	19,04	0,0000
COD_TARIF2-Passeio Importado	3,6590	0,1497	24,44	0,0000
COD_TARIF3-Pick-up	2,3901	0,0951	25,13	0,0000
COD_TARIF4-Moto	2,9297	0,4420	6,63	0,0000
REGIAOMG-Int	-0,8474	0,2028	-4,18	0,0000
REGIAOMG_RM	-0,8309	0,1990	-4,18	0,0000
REGIAORJ-Int	0,1157	0,2169	0,53	0,5943
REGIAORJ_RM	-1,1471	0,1915	-5,99	0,0000
REGIAOSP-int	-2,2514	0,1871	-12,04	0,0000
REGIAOSP_RM	-2,3409	0,1862	-12,57	0,0000
IDADE26 a 35 anos	-1,9211	0,2050	-9,37	0,0000
IDADE36 a 45 anos	-2,6566	0,1983	-13,39	0,0000
IDADE46 a 55 anos	-2,5536	0,2024	-12,62	0,0000
IDADE56 anos e +	-2,9848	0,1979	-15,08	0,0000
ANO_MODELO2011-2013	-0,0400	0,1113	-0,36	0,7195
ANO_MODELO2014-2017	-0,1054	0,1121	-0,94	0,3478
ANO_MODELO2018-2021	0,4207	0,1476	2,85	0,0047
SEXOM	-0,2705	0,0875	-3,09	0,0022

Novamente, para interpretar os coeficientes do modelo Gama para severidade semestral de sinistros da cobertura incêndios, faremos o uso das relatividades. Como a variável resposta não está transformada, a relatividade é calculada utilizando a função inversa da função de ligação usada no modelo. No caso em estudo, foi aplicado a função de ligação *log*, consequentemente, a relatividade é calculada fazendo o exponencial do coeficiente. Os resultados encontrados estão na Tabela 4.5.

Mais uma vez, usando a covariável ano modelo para exemplificar a interpretação, vemos na tabela da Tabela 4.5 que a média da classe Ano Modelo 2011-2013 é aproximadamente 4% menor do que a média do *baseline*. Além disso, a média da classe Ano Modelo 2014-2017 é aproximadamente 10% menor do que a média da classe base e, a Classe Ano Modelo 2018-2021 apresentou média 52% maior que a categoria base.

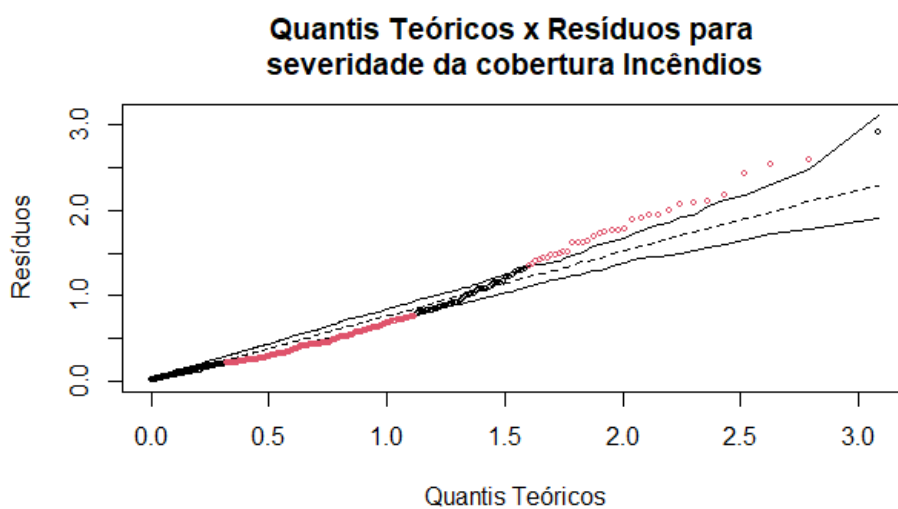


Figura 4.43 – Hnp para Severidade da cobertura Incêndios

Tabela 4.5 – Relatividades para Severidade da cobertura Incêndios

Fator de Risco	Classe	Severidade Relatividade
Região	ES	1,0000
	MG - Int	0,4285
	MG - RM	0,4357
	RJ - Int	1,1226
	RJ - RM	0,3176
	SP - Int	0,1052
	SP - RM	0,0962
Sexo	F	1,0000
	M	0,7630
Idade	18-25	1,0000
	26-35	0,1464
	36-45	0,0702
	46-55	0,0778
	55+	0,0506
Código Tarifário	Passeio Nacional	1,0000
	Passeio Importado	38,8210
	Pick-up	10,9151
	Moto	18,7221
Ano Modelo	1990-2010	1,0000
	2011-2013	0,9608
	2014-2017	0,8999
	2018-2021	1,5231

4.3 Ajuste e interpretação dos modelos Tweedie

O modelo Tweedie é construído a partir do parâmetro de dispersão dos modelos usando a distribuição Gama apresentados anteriormente (cada um para seu respectivo tipo de cobertura). Além disso, frequência zero também é incluída no modelo devido ao fato de que a Tweedie tem, como definido na Seção 3.5, uma massa de probabilidade positiva no zero, representando os casos sem indenizações, ou seja, frequência igual a zero ou casos em que a seguradora não pagou indenização para o sinistro (devido à franquia).

De modo geral, para quase todos os tipos de cobertura, os modelos Tweedie apresentaram todas as covariáveis significativas, com exceção do modelo para cobertura incêndios. A avaliação do ajuste será apenas utilizando a *Deviance* que, para todos os tipos de cobertura, foi menor do que os graus de liberdade do modelo, que é a média da distribuição Qui-Quadrado, o que, conseqüentemente torna válida a utilização do modelo usando a Tweedie.

No caso da cobertura incêndios, duas covariáveis não foram significativas: sexo e região. Sendo assim, em primeiro lugar, foi retirada a variável explicativa sexo do modelo e realizado um novo ajuste. Nesse novo ajuste, a região permanecia como não significativa e, portanto, foi realizado um novo ajuste sem sexo e região. Entretanto, ao retirar a região da modelagem, apenas o código tarifário fica significativo e o intercepto apresenta valor-p maior do que 0,05, ou seja, não significativo. Desta forma, optamos por manter a variável explicativa região no modelo. A Tabela 4.6 mostra os coeficientes estimados e seus respectivos valores-p para o modelo da cobertura incêndios. Para as demais coberturas, os modelos Tweedie se encontram no Anexo C.

Tabela 4.6 – Resultado do ajuste do modelo Tweedie para cobertura Incêndios

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2,1668	0,6756	3,21	0,0014
COD_TARIF2-Passeio Importado	1,6979	0,3608	4,71	0,0000
COD_TARIF3-Pick-up	1,8445	0,2763	6,67	0,0000
COD_TARIF4-Moto	-2,2894	1,0632	-2,15	0,0315
REGIAOMG-Int	-0,0973	0,5776	-0,17	0,8663
REGIAOMG_RM	-0,0108	0,5569	-0,02	0,9846
REGIAORJ-Int	0,4817	0,5957	0,81	0,4189
REGIAORJ-RM	0,0362	0,5546	0,07	0,9479
REGIAOSP-int	-0,8677	0,5380	-1,61	0,1071
REGIAOSP-RM	-0,6506	0,5216	-1,25	0,2125
IDADE26 a 35 anos	-1,0155	0,5218	-1,95	0,0519
IDADE36 a 45 anos	-1,1220	0,4972	-2,26	0,0242
IDADE46 a 55 anos	-1,3475	0,5104	-2,64	0,0084
IDADE56 anos e +	-1,4636	0,4987	-2,93	0,0034
ANO_MODELO2011-2013	-0,0647	0,3534	-0,18	0,8548
ANO_MODELO2014-2017	-0,1806	0,3407	-0,53	0,5960
ANO_MODELO2018-2021	-0,7513	0,4059	-1,85	0,0644

Como pode ser visto na Tabela 4.6, todas as categorias de código tarifário apresentaram valor-p menor do que 0,05. O mesmo comportamento pode ser observado para a covariável idade (a categoria 26 a 35 anos apresentou um valor muito próximo de 0,05). Para ano do modelo, a categoria 2018-2021 apresentou um valor- de aproximadamente 0,06, o que não permite a gente falar que essa categoria é não significativa devido à sua proximidade com o nível que está sendo considerado (5%). As demais faixas de Ano do Modelo não são significativas. Por fim, podemos notar que todas as categorias da variável explicativa região apresentaram valor-p acima de 0,05, mas, devido ao explicado anteriormente, ela não será retirada do modelo.

Da mesma forma que no modelo Gama, a relatividade da Tweedie é calculada tomando o exponencial do coeficiente pois a função de ligação utilizada no ajuste do modelo Tweedie é a *Log*. Para exemplificar a interpretação das relatividades apresentadas na Tabela 4.7, podemos ver que a média da classe Ano Modelo 2011-2013 é aproximadamente 6% menor do que a média do *baseline*. Além disso, a média da classe região RJ - Interior é aproximadamente 62% maior que a região ES, que é a categoria base desta variável.

Tabela 4.7 – Relatividades do modelo Tweedie para cobertura Incêndios

Fator de Risco	Classe	Relatividade
Região	ES	1,0000
	MG - Int	0,9073
	MG - RM	0,9893
	RJ - Int	1,6188
	RJ - RM	1,0369
	SP - Int	0,4199
	SP - RM	0,5217
Sexo	F	1,0000
	M	1,0000
Idade	18-25	1,0000
	26-35	0,3622
	36-45	0,3256
	46-55	0,2599
	55+	0,2314
Código Tarifário	Passeio Nacional	1,0000
	Passeio Importado	5,4622
	Pick-up	6,3247
	Moto	0,1013
Ano Modelo	1990-2010	1,0000
	2011-2013	0,9374
	2014-2017	0,8347
	2018-2021	0,4717

4.4 Prêmio de Risco

Nessa seção serão apresentados valores estimados do Prêmio de Risco para algumas células tarifárias. Sendo assim, selecionamos para a cobertura incêndios uma região (Minas Gerais - Região Metropolitana) e um código tarifário (Passeio Nacional), e com essas variáveis fixas, faremos comparações quanto a sexo, idade e ano do modelo. Ademais será feita uma comparação entre os prêmios encontrados usando modelo multiplicativo e usando modelo Tweedie. Devido à grande quantidade de coberturas e de células tarifárias, os demais valores estimados não serão expostos nesse trabalho.

Vale destacar que, como dito na Seção 4.2, o ajuste do modelo da severidade foi realizado com o *log* da exposição como *offset*. Desta forma, quando multiplicamos os valores estimados para as células tarifárias para frequência pelos valores encontrados para a severidade, a frequência está sendo multiplicada duas vezes. Sendo assim, para estimar o Prêmio de Risco basta fazer o modelo multiplicativo para a severidade.

Na Tabela 4.8, encontram-se os Prêmios de Riscos estimados usando modelo multiplicativo para cobertura incêndios em algumas células tarifárias de Minas Gerais - Região Metropolitana e código tarifário Passeio Nacional.

Tabela 4.8 – Prêmio de Risco estimado para cobertura Incêndios usando modelo multiplicativo

Prêmio de Risco para cobertura Incêndios usando Modelo Multiplicativo									
Região: Minas Gerais – Região Metropolitana e Código Tarifário: Passeio Nacional									
Faixa de idade	Ano do modelo 1990-2010		Ano do modelo 2011-2013		Ano do modelo 2014-2017		Ano do modelo 2018-2021		
	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	
18-25	37,08	48,59	35,62	46,69	33,37	43,73	56,47	74,01	
26-35	5,43	7,12	5,22	6,84	4,89	6,40	8,27	10,84	
36-45	2,60	3,41	2,50	3,28	2,34	3,07	3,96	5,19	
46-55	2,88	3,78	2,77	3,63	2,60	3,40	4,39	5,76	
56 e +	1,87	2,46	1,80	2,36	1,69	2,21	2,85	3,74	

Com os resultados expostos na Tabela 4.8, podemos notar que o Prêmio de Risco estimado via modelo multiplicativo para a cobertura Incêndios é sempre maior para o sexo feminino. Ademais, quanto mais velho o condutor, menor é o Prêmio de Risco, sendo que o valor sofre uma queda de aproximadamente 85% quando a idade passa da faixa 18 a 25 anos para a faixa 26 a 35 anos, e dessa faixa adiante, a diferença entre os prêmios se torna mais suave. Por último, o Prêmio de Risco apresentou uma queda de 4% quando passa do ano do modelo 1990-2010 para 2011-2013; de 2011-2013 para 2014-2017 a queda no valor do prêmio foi de aproximadamente 6% e, de 2014-2017 para 2018-2021 o Prêmio de Risco estimado aumentou aproximadamente 66%. Ou seja, a princípio temos uma redução no valor do prêmio conforme a redução da idade do veículo, mas no grupo de idade mais recente, o Prêmio de Risco aumenta consideravelmente.

As alterações nos valores dos Prêmios de Riscos quando mudamos de células tarifárias citadas anteriormente poderão ser vistas claramente nos gráficos da Figura 4.44

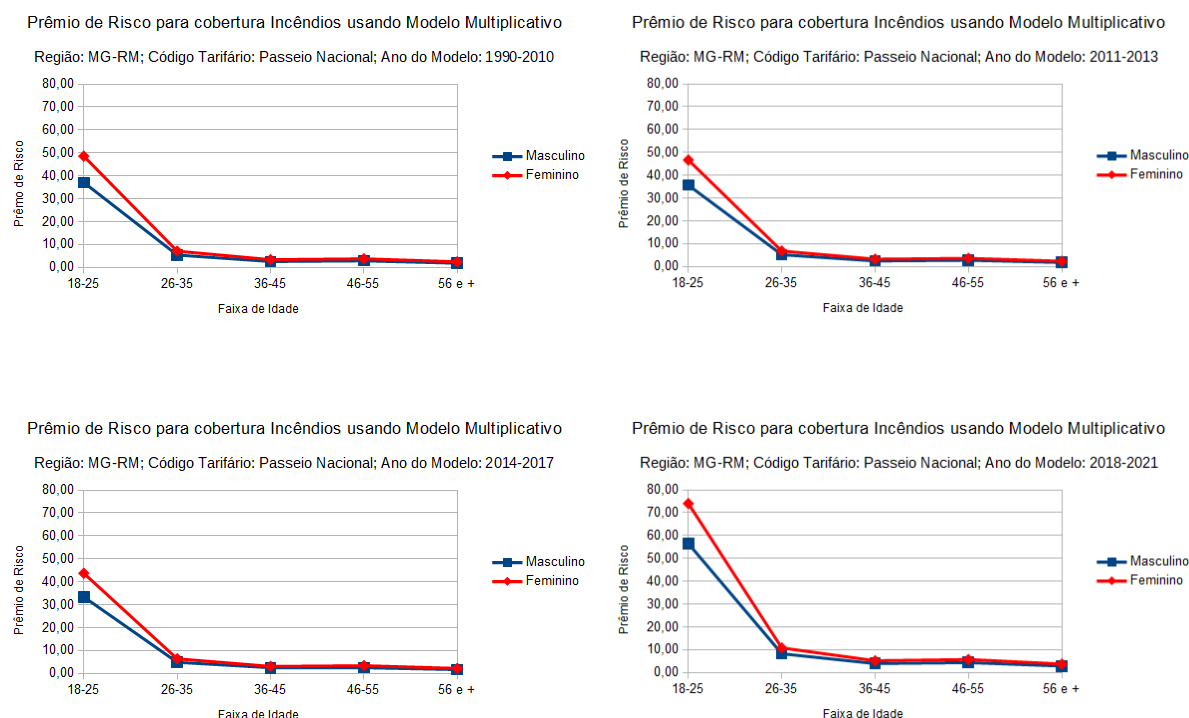


Figura 4.44 – Gráficos com Prêmio de Risco estimado para cobertura Incêndios usando modelo multiplicativo

Na Tabela 4.9, encontram-se os Prêmios de Riscos estimados usando modelo Tweedie para cobertura incêndios em algumas células tarifárias de Minas Gerais - Região Metropolitana e código tarifário Passeio Nacional.

Tabela 4.9 – Prêmio de Risco estimado para cobertura Incêndios usando modelo Tweedie

Prêmio de Risco para cobertura Incêndios usando Modelo Tweedie								
Região: Minas Gerais – Região Metropolitana e Código Tarifário: Passeio Nacional								
Faixa de idade	Ano do modelo 1990-2010		Ano do modelo 2011-2013		Ano do modelo 2014-2017		Ano do modelo 2018-2021	
	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino
18-25	8,64	8,64	8,10	8,10	7,21	7,21	4,07	4,07
26-35	3,13	3,13	2,93	2,93	2,61	2,61	1,48	1,48
36-45	2,81	2,81	2,64	2,64	2,35	2,35	1,33	1,33
46-55	2,24	2,24	2,10	2,10	1,87	1,87	1,06	1,06
56 e +	2,00	2,00	1,87	1,87	1,67	1,67	0,94	0,94

Com os resultados expostos na Tabela 4.8, podemos notar que o Prêmio de Risco estimado via modelo Tweedie para a cobertura Incêndios é o mesmo tanto para o sexo masculino quanto para o feminino. Isso é justificado pelo fato da relatividade do sexo ser 1. Ademais, quanto mais velho o condutor, menor é o Prêmio de Risco, sendo que o valor sofre uma queda de aproximadamente 64% quando a idade passa da faixa 18 a 25 anos para a faixa 26 a 35 anos, e dessa faixa adiante, a diferença entre os prêmios se torna mais suave.

Por último, o Prêmio de Risco apresentou uma queda de aproximadamente 6% quando passa do ano do modelo 1990-2010 para 2011-2013; de 2011-2013 para 2014-2017 a queda no valor do prêmio foi de aproximadamente 1% e, de 2014-2017 para 2018-2021 a queda foi de aproximadamente 43%. Ou seja, quanto mais novo é o veículo de código tarifário Passeio Nacional, menor o valor do Prêmio de Risco estimado pelo modelo Tweedie.

As alterações nos valores dos Prêmios de Riscos quando mudamos de células tarifárias citadas anteriormente poderão ser vistas claramente nos gráficos da Figura 4.45

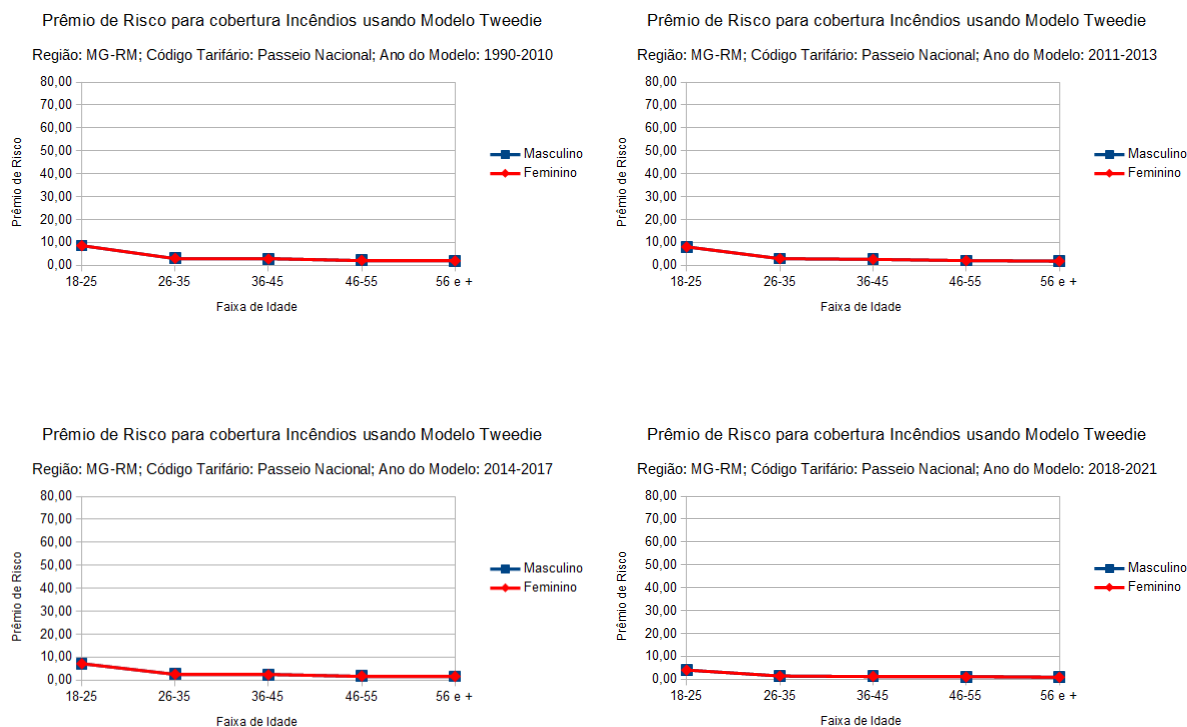


Figura 4.45 – Gráficos com Prêmio de Risco estimado para cobertura Incêndios usando modelo Tweedie

Finalmente, fazendo uma comparação entre os Prêmios de Riscos estimados usando modelo multiplicativo (apresentado na Tabela 4.8 e na Figura 4.44) e usando modelo Tweedie (apresentado na Tabela 4.9 e na Figura 4.45), podemos notar algumas diferenças. Primeiramente, o valor estimado para o primeiro grupo de idade do condutor do sexo masculino usando o modelo multiplicativo é aproximadamente 77% maior do que o estimado usando o modelo Tweedie, e para o para o primeiro grupo de idade do condutor do sexo feminino, essa diferença é de 83%. Para as demais categorias de idade, os valores estimados pelos dois modelos são relativamente próximos.

Em segundo lugar, nota-se que, enquanto no modelo multiplicativo o Prêmio de Risco para o sexo feminino é maior do que para o sexo masculino, no modelo Tweedie não tem essa distinção por sexo devido ao fato dessa variável não estar presente no ajuste final do modelo, fazendo com que o coeficiente estimado seja zero. Outro ponto a destacar é que no modelo multiplicativo temos aumento do prêmio no último grupo de ano do modelo do

veículo, enquanto no Tweedie o Prêmio de Risco seguiu o padrão de queda apresentado pelos grupos de veículos mais velhos. Por fim, de modo geral, o Prêmio de Risco estimado pelo modelo Tweedie é menor do que o estimado através do modelo multiplicativo. Isso é explicado pelo fato de que o modelo Tweedie inclui indenizações e frequências iguais a zero, ao passo de que no modelo multiplicativo elas não estão incluídas, fazendo com que o Prêmio de Risco estimado pelo modelo Tweedie seja menor do que o estimado pelo modelo multiplicativo.

5 Conclusão

Todos os gráficos e tabelas apresentados nesse trabalho são de autoria própria e foram construídos usando os *softwares* R e *Microsoft Excel* a partir dos dados obtidos do sistema da Susep.

Uma vez concluída a exposição dos resultados, pode-se dizer que os objetivos desse trabalho foram cumpridos. Com a análise descritiva da base de dados, pôde-se identificar alguns comportamentos que seriam esperados para os valores estimados para o Prêmio de Risco de cada célula tarifária. Boa parte desses comportamentos foram observados nos resultados finais, fato esse que facilitou a interpretação desses resultados.

Além disso, foram testados e comparados dois métodos de estimação do prêmio: modelo multiplicando frequência vezes severidade, com tentativas de ajustes usando as distribuições Poisson, Binomial Negativa e Log-Normal para a frequência e as distribuições Gama e Log-Normal para a severidade, e o modelo de mistura Tweedie, onde a frequência e a severidade são ajustadas conjuntamente.

Entretanto, o ajuste desses modelos ficou prejudicado devido à grande presença de frequência de sinistros igual a zero. A saída para tal problema utilizada no presente estudo foi retirar as frequências iguais a zero no momento de fazer o ajuste para cada tipo de categorias. Tal procedimento foi realizado para melhorar a estimativa dos coeficientes dos modelos. Ainda assim, os valores estimados para o Prêmio de Risco estão abaixo do praticado no mercado. São dois os motivos para isso: primeiramente, os seguros de veículos, em sua maioria, são feitos para o período de um ano, conseqüentemente, seus prêmios são calculados para o ano e, a base de dados usada nesse trabalho é semestral; em segundo lugar, os prêmios estimados por esse trabalho estão separados por tipo de cobertura, enquanto no mercado não tem essa distinção.

Desta forma, um ponto que pode ser melhorado em trabalhos futuros nesse processo para estimar o Prêmio de Risco é testar a utilização de modelos que usam distribuições com zeros inflacionados ou até mesmo outras distribuições de mistura que podem se ajustar melhor à base de dados.

6 Referências

- Amarante, C. e Ferreira, P. **Aspectos Atuariais e Contábeis das Provisões Técnicas**. Escola Nacional de Seguros. Rio de Janeiro, 2009.
- Bastos, S. F. **Teorema de Wilks: uma abordagem Teórica e Assintótica via Simulação de Monte Carlo** 2014. Monografia de Estatística, Departamento de Estatística, Universidade Federal Fluminense, Rio de Janeiro, 2014.
- Bolfarine, H.; Sandoval, M. C. **Introdução à Inferência Estatística** 2000.
- Charpentier, A. **Computational actuarial science with R**. CRC Press.(Ed.2014).
- De Jong, Piet; Gillian Z. Heller **Generalized linear models for insurance data**. Cambridge: Cambridge University Press, 2008.
- Denuit, Michel, et al. **Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems**. John Wiley e Sons, 2007.
- Escola Nacional de Seguros. **Conceitos básicos de seguros**. Supervisão e coordenação metodológica da Diretoria de Ensino Técnico. Assessoria técnica de José Antonio Menezes Varanda. – 7. ed. – Rio de Janeiro: Funenseg, 2011
- Farias, T.A. ; de Jesus, J.C. **Tarifação De Seguros De Automóveis No Mercado Do Nordeste**. IX Simpósio de Atuária. Recife, 2020.
- Ferreira, Paulo Pereira. **Modelos de precificação e ruína para seguros de curto prazo**. Rio de Janeiro: Funenseg, Reimpressão:2002.
- Jewell, W.S. **The credible distribution**. Astin Bulletin, Leuven, v.7, n.3, pp. 237-269, 1974.
- McCullagh, P.; Nelder, J.A. **Generalized Linear Models**. Mathematical Statistics of Generalized Linear Models. London: Chapman and Hall, 1989.
- Montgomery, D. C.;Runger, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**, sexta edição, 2016.
- Moral, R.A.; Hinde J, Demétrio C.G.B. **Half-Normal Plots and Overdispersed Models in R: The hnp Package**. Journal of Statistical Software. 2017, 1-23.
- Morgado, W. L. **Método de Classificação de Risco Aplicado ao Seguro Automóvel**. 2004. 105 f. Dissertação de Mestrado em Engenharia Elétrica, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2004.
- Nelder, J.A.; Wedderburn, R.W.M. **Generalized Linear Models**. Journal of the Royal Statistical Society. Series A (General). Vol. 135, No. 3 (1972), p. 370-384. Wiley, 1972.
- Ohlsson, E; Johansson, **Non-life insurance pricing with generalized linear Models**. Berlin: Springer, 2010.

Pacheco, Ricardo. **Matemática Atuarial de Seguros de Danos**. Editora Atlas, 2014.

Quijano Xacur, O.A.; Garrido, J. **Generalised linear models for aggregate claims: to Tweedie or not?**. Eur. Actuar. J. 5, 181–202 (2015).

Silva, Y.R; Afonso, L.E. **A Comparative Study of Pricing Methods of Automobile Insurance in Brazil**. Revista Brasileira de Risco e Seguro. 2015, 25-44.

Silvério, A.F. **Modelagem Gln Aplicada À Atuária: Uma Utilização Dos Modelos Lineares Generalizados Na Precificação De Seguros**. 2015. Monografia de Graduação em Ciências Atuariais, Departamento de Administração, Universidade Federal do Ceará. Fortaleza, 2015.

Souza, Antonio Lober Ferreira et al. **Dicionário de seguros: vocabulário conceituado de seguros**. Rio de Janeiro: IRB-Brasil Re S.A./Funenseg, 2000.

SUSEP. **História do Seguro**. Disponível em <http://www.susep.gov.br/menu/a-susep/historia-do-seguro> Acesso em: 20 de junho de 2021

SUSEP. **Síntese Mensal Outubro 2021**. Disponível em <http://novosite.susep.gov.br/noticias/susep-divulga-sintese-mensal-com-dados-do-setor-em-outubro-2/> Acesso em: 7 de janeiro de 2022

Turkman, M. A. A.; Silva, G. L. **Modelos Lineares Generalizados: da teoria à prática**. Lisboa: Universidade Técnica de Lisboa, 2000.

ANEXO A – Modelos para Frequência

Tabela A.1 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Roubo/Furto

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4,6676	0,0931	-50,14	0,0000
COD_TARIF2-Passeio Importado	0,3221	0,0584	5,52	0,0000
COD_TARIF3-Pick-up	0,3892	0,0512	7,59	0,0000
COD_TARIF4-Moto	1,0888	0,0562	19,39	0,0000
REGIAOMG-Int	-0,3206	0,0812	-3,95	0,0001
REGIAOMG_RM	0,0612	0,0786	0,78	0,4367
REGIAORJ-Int	0,4622	0,0820	5,64	0,0000
REGIAORJ_RM	0,9616	0,0784	12,26	0,0000
REGIAOSP-int	0,0073	0,0772	0,09	0,9245
REGIAOSP_RM	0,6987	0,0761	9,18	0,0000
IDADE26 a 35 anos	-0,2897	0,0673	-4,30	0,0000
IDADE36 a 45 anos	-0,5858	0,0678	-8,65	0,0000
IDADE46 a 55 anos	-0,8107	0,0683	-11,87	0,0000
IDADE56 anos e +	-1,0660	0,0693	-15,39	0,0000
ANO_MODELO2011-2013	-0,1673	0,0569	-2,94	0,0034
ANO_MODELO2014-2017	-0,3332	0,0563	-5,91	0,0000
ANO_MODELO2018-2021	-0,5074	0,0582	-8,72	0,0000
SEXOM	0,1956	0,0398	4,91	0,0000

Tabela A.2 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Colisão Parcial

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2,9799	0,0516	-57,71	0,0000
COD_TARIF2-Passeio Importado	-0,1544	0,0326	-4,74	0,0000
COD_TARIF3-Pick-up	0,0810	0,0319	2,54	0,0113
COD_TARIF4-Moto	-1,2263	0,0349	-35,17	0,0000
REGIAOMG-Int	0,1229	0,0450	2,73	0,0064
REGIAOMG_RM	0,0365	0,0448	0,82	0,4152
REGIAORJ-Int	-0,2518	0,0464	-5,43	0,0000
REGIAORJ_RM	-0,3619	0,0458	-7,90	0,0000
REGIAOSP-int	-0,0277	0,0446	-0,62	0,5341
REGIAOSP_RM	-0,1083	0,0447	-2,42	0,0156
IDADE26 a 35 anos	-0,3423	0,0389	-8,81	0,0000
IDADE36 a 45 anos	-0,4052	0,0388	-10,46	0,0000
IDADE46 a 55 anos	-0,4534	0,0391	-11,59	0,0000
IDADE56 anos e +	-0,4385	0,0392	-11,19	0,0000
ANO_MODELO2011-2013	0,3539	0,0349	10,15	0,0000
ANO_MODELO2014-2017	0,6622	0,0342	19,34	0,0000
ANO_MODELO2018-2021	0,9626	0,0343	28,03	0,0000

Tabela A.3 – Resultado do ajuste do modelo Log-Normal para Frequência da cobertura Colisão Perda Total

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3,8220	0,0693	-55,19	0,0000
COD_TARIF2-Passeio Importado	0,6636	0,0433	15,32	0,0000
COD_TARIF3-Pick-up	0,2402	0,0410	5,85	0,0000
COD_TARIF4-Moto	0,3145	0,0451	6,97	0,0000
REGIAOMG-Int	-0,0091	0,0598	-0,15	0,8787
REGIAOMG_RM	-0,0737	0,0591	-1,25	0,2133
REGIAORJ-Int	0,1903	0,0619	3,07	0,0022
REGIAORJ_RM	-0,0578	0,0599	-0,96	0,3348
REGIAOSP-int	-0,2516	0,0577	-4,36	0,0000
REGIAOSP_RM	-0,5007	0,0582	-8,60	0,0000
IDADE26 a 35 anos	-0,6067	0,0524	-11,59	0,0000
IDADE36 a 45 anos	-0,7711	0,0522	-14,78	0,0000
IDADE46 a 55 anos	-0,8534	0,0527	-16,19	0,0000
IDADE56 anos e +	-0,9651	0,0532	-18,14	0,0000
ANO_MODELO2011-2013	-0,2164	0,0444	-4,87	0,0000
ANO_MODELO2014-2017	-0,5972	0,0441	-13,54	0,0000
ANO_MODELO2018-2021	-0,8788	0,0457	-19,23	0,0000
SEXOM	0,1671	0,0313	5,34	0,0000

Tabela A.4 – Resultado do ajuste do modelo Log-Normal para Frequência de outras coberturas

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1,2172	0,0539	-22,58	0,0000
COD_TARIF2-Passeio Importado	-0,1955	0,0351	-5,56	0,0000
COD_TARIF3-Pick-up	-0,0045	0,0346	-0,13	0,8962
COD_TARIF4-Moto	-1,3007	0,0353	-36,83	0,0000
REGIAOMG-Int	-0,1008	0,0467	-2,16	0,0311
REGIAOMG_RM	0,0884	0,0469	1,89	0,0597
REGIAORJ-Int	0,2141	0,0479	4,47	0,0000
REGIAORJ_RM	0,3583	0,0470	7,63	0,0000
REGIAOSP-int	0,0402	0,0465	0,86	0,3875
REGIAOSP_RM	0,6387	0,0467	13,68	0,0000
IDADE26 a 35 anos	0,1284	0,0403	3,19	0,0015
IDADE36 a 45 anos	0,0702	0,0403	1,74	0,0815
IDADE46 a 55 anos	0,0145	0,0403	0,36	0,7184
IDADE56 anos e +	0,0371	0,0405	0,92	0,3594
ANO_MODELO2011-2013	-0,1317	0,0355	-3,71	0,0002
ANO_MODELO2014-2017	-0,3455	0,0352	-9,82	0,0000
ANO_MODELO2018-2021	-0,8489	0,0357	-23,81	0,0000
SEXOM	-0,1052	0,0249	-4,22	0,0000

ANEXO B – Modelos para Severidade

Tabela B.1 – Resultado do ajuste do modelo Gama para Severidade da cobertura Roubo/Furto

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4,9212	0,1690	29,12	0,0000
COD_TARIF2-Passeio Importado	3,6485	0,1060	34,44	0,0000
COD_TARIF3-Pick-up	2,4681	0,0930	26,52	0,0000
COD_TARIF4-Moto	2,2853	0,1018	22,46	0,0000
REGIAOMG-Int	-0,5942	0,1475	-4,03	0,0001
REGIAOMG-RM	-0,8918	0,1428	-6,25	0,0000
REGIAORJ-Int	0,1083	0,1488	0,73	0,4670
REGIAORJ-RM	-0,7352	0,1422	-5,17	0,0000
REGIAOSP-int	-2,1787	0,1402	-15,54	0,0000
REGIAOSP-RM	-1,9158	0,1383	-13,86	0,0000
IDADE26 a 35 anos	-1,8971	0,1223	-15,52	0,0000
IDADE36 a 45 anos	-2,4482	0,1230	-19,90	0,0000
IDADE46 a 55 anos	-2,3357	0,1240	-18,84	0,0000
IDADE56 anos e +	-2,3349	0,1256	-18,60	0,0000
ANO_MODELO2011-2013	-0,2565	0,1032	-2,49	0,0131
ANO_MODELO2014-2017	-0,4706	0,1023	-4,60	0,0000
ANO_MODELO2018-2021	0,2539	0,1057	2,40	0,0165
SEXOM	-0,4657	0,0723	-6,44	0,0000

Tabela B.2 – Resultado do ajuste do modelo Gama para Severidade da cobertura Colisão Parcial

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3,8410	0,1634	23,51	0,0000
COD_TARIF2-Passeio Importado	3,9639	0,1010	39,25	0,0000
COD_TARIF3-Pick-up	2,4359	0,0989	24,63	0,0000
COD_TARIF4-Moto	3,3331	0,1075	31,01	0,0000
REGIAOMG-Int	-0,9225	0,1384	-6,66	0,0000
REGIAOMG-RM	-0,8814	0,1382	-6,38	0,0000
REGIAORJ-Int	-0,0524	0,1431	-0,37	0,7141
REGIAORJ-RM	-0,8084	0,1415	-5,71	0,0000
REGIAOSP-int	-2,3694	0,1378	-17,19	0,0000
REGIAOSP-RM	-2,0292	0,1378	-14,72	0,0000
IDADE26 a 35 anos	-2,2354	0,1202	-18,60	0,0000
IDADE36 a 45 anos	-2,8320	0,1198	-23,64	0,0000
IDADE46 a 55 anos	-2,7350	0,1207	-22,65	0,0000
IDADE56 anos e +	-2,6694	0,1208	-22,09	0,0000
ANO_MODELO2011-2013	-0,3766	0,1075	-3,50	0,0005
ANO_MODELO2014-2017	-0,8501	0,1058	-8,04	0,0000
ANO_MODELO2018-2021	-0,2063	0,1059	-1,95	0,0518
SEXOM	-0,3546	0,0732	-4,84	0,0000

Tabela B.3 – Resultado do ajuste do modelo Gama para Severidade da cobertura Colisão Perda Total

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5,1920	0,1928	26,94	0,0000
COD_TARIF2-Passeio Importado	3,7414	0,1199	31,21	0,0000
COD_TARIF3-Pick-up	2,3892	0,1143	20,90	0,0000
COD_TARIF4-Moto	2,4659	0,1258	19,61	0,0000
REGIAOMG-Int	-0,7482	0,1662	-4,50	0,0000
REGIAOMG_RM	-0,9559	0,1650	-5,79	0,0000
REGIAORJ-Int	0,0017	0,1721	0,01	0,9920
REGIAORJ_RM	-0,7214	0,1666	-4,33	0,0000
REGIAOSP-int	-2,0817	0,1607	-12,96	0,0000
REGIAOSP_RM	-2,0628	0,1619	-12,74	0,0000
IDADE26 a 35 anos	-1,9814	0,1452	-13,65	0,0000
IDADE36 a 45 anos	-2,5285	0,1444	-17,51	0,0000
IDADE46 a 55 anos	-2,3031	0,1460	-15,77	0,0000
IDADE56 anos e +	-2,5277	0,1476	-17,13	0,0000
ANO_MODELO2011-2013	-0,4952	0,1237	-4,00	0,0001
ANO_MODELO2014-2017	-0,6083	0,1227	-4,96	0,0000
ANO_MODELO2018-2021	0,3237	0,1267	2,56	0,0108
SEXOM	-0,3506	0,0870	-4,03	0,0001

Tabela B.4 – Resultado do ajuste do modelo Gama para Severidade de outras coberturas

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2,0096	0,1878	10,70	0,0000
COD_TARIF2-Passeio Importado	3,7182	0,1220	30,47	0,0000
COD_TARIF3-Pick-up	2,0602	0,1205	17,09	0,0000
COD_TARIF4-Moto	3,2355	0,1232	26,26	0,0000
REGIAOMG-Int	-0,9549	0,1622	-5,89	0,0000
REGIAOMG_RM	-1,2891	0,1627	-7,92	0,0000
REGIAORJ-Int	-0,0844	0,1665	-0,51	0,6123
REGIAORJ_RM	-0,9243	0,1635	-5,65	0,0000
REGIAOSP-int	-2,5210	0,1615	-15,61	0,0000
REGIAOSP_RM	-2,8340	0,1618	-17,52	0,0000
IDADE26 a 35 anos	-1,9518	0,1401	-13,93	0,0000
IDADE36 a 45 anos	-2,6100	0,1397	-18,69	0,0000
IDADE46 a 55 anos	-2,5924	0,1397	-18,56	0,0000
IDADE56 anos e +	-2,6436	0,1404	-18,83	0,0000
ANO_MODELO2011-2013	-0,5499	0,1239	-4,44	0,0000
ANO_MODELO2014-2017	-1,1452	0,1227	-9,33	0,0000
ANO_MODELO2018-2021	-0,3322	0,1237	-2,68	0,0074
SEXOM	-0,4261	0,0868	-4,91	0,0000

ANEXO C – Modelos Tweedie

Tabela C.1 – Resultado do ajuste do modelo Tweedie para cobertura Roubo/Furto

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3,9673	0,1901	20,87	0,0000
COD_TARIF2-Passeio Importado	3,0928	0,1114	27,77	0,0000
COD_TARIF3-Pick-up	2,2675	0,1049	21,61	0,0000
COD_TARIF4-Moto	1,4652	0,1343	10,91	0,0000
REGIAOMG-Int	-0,4302	0,1588	-2,71	0,0068
REGIAOMG_RM	-0,6175	0,1576	-3,92	0,0001
REGIAORJ-Int	0,3224	0,1603	2,01	0,0446
REGIAORJ_RM	-0,3938	0,1549	-2,54	0,0112
REGIAOSP-int	-1,7109	0,1557	-10,99	0,0000
REGIAOSP_RM	-1,6276	0,1516	-10,73	0,0000
IDADE26 a 35 anos	-1,2862	0,1357	-9,48	0,0000
IDADE36 a 45 anos	-1,9199	0,1363	-14,08	0,0000
IDADE46 a 55 anos	-1,8461	0,1373	-13,44	0,0000
IDADE56 anos e +	-2,0895	0,1371	-15,24	0,0000
ANO_MODELO2011-2013	-0,1133	0,1226	-0,92	0,3559
ANO_MODELO2014-2017	-0,2515	0,1192	-2,11	0,0351
ANO_MODELO2018-2021	0,3059	0,1190	2,57	0,0103
SEXOM	-0,2011	0,0808	-2,49	0,0129

Tabela C.2 – Resultado do ajuste do modelo Tweedie para cobertura Colisão Parcial

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3,2854	0,1952	16,83	0,0000
COD_TARIF2-Passeio Importado	3,7731	0,1230	30,67	0,0000
COD_TARIF3-Pick-up	2,3374	0,1243	18,81	0,0000
COD_TARIF4-Moto	2,5508	0,1370	18,62	0,0000
REGIAOMG-Int	-0,7794	0,1673	-4,66	0,0000
REGIAOMG_RM	-0,8379	0,1637	-5,12	0,0000
REGIAORJ-Int	-0,0291	0,1692	-0,17	0,8634
REGIAORJ_RM	-0,7613	0,1640	-4,64	0,0000
REGIAOSP-int	-2,1984	0,1660	-13,24	0,0000
REGIAOSP_RM	-1,9748	0,1594	-12,39	0,0000
IDADE26 a 35 anos	-1,8283	0,1407	-13,00	0,0000
IDADE36 a 45 anos	-2,3685	0,1405	-16,85	0,0000
IDADE46 a 55 anos	-2,3219	0,1423	-16,31	0,0000
IDADE56 anos e +	-2,4469	0,1411	-17,35	0,0000
ANO_MODELO2011-2013	-0,3395	0,1282	-2,65	0,0082
ANO_MODELO2014-2017	-0,6848	0,1271	-5,39	0,0000
ANO_MODELO2018-2021	-0,1475	0,1267	-1,16	0,2444
SEXOM	-0,2505	0,0881	-2,84	0,0046

Tabela C.3 – Resultado do ajuste do modelo Tweedie para cobertura Colisão Perda Total

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4,4129	0,2243	19,68	0,0000
COD_TARIF2-Passeio Importado	3,3857	0,1354	25,01	0,0000
COD_TARIF3-Pick-up	2,1859	0,1332	16,41	0,0000
COD_TARIF4-Moto	1,6776	0,1572	10,67	0,0000
REGIAOMG-Int	-0,5962	0,1891	-3,15	0,0017
REGIAOMG-RM	-0,7570	0,1874	-4,04	0,0001
REGIAORJ-Int	0,0164	0,1931	0,08	0,9324
REGIAORJ-RM	-0,6788	0,1872	-3,63	0,0003
REGIAOSP-int	-1,7888	0,1844	-9,70	0,0000
REGIAOSP-RM	-1,9522	0,1841	-10,60	0,0000
IDADE26 a 35 anos	-1,4759	0,1635	-9,03	0,0000
IDADE36 a 45 anos	-1,9784	0,1626	-12,17	0,0000
IDADE46 a 55 anos	-1,8762	0,1633	-11,49	0,0000
IDADE56 anos e +	-2,1382	0,1637	-13,06	0,0000
ANO_MODELO2011-2013	-0,2495	0,1496	-1,67	0,0957
ANO_MODELO2014-2017	-0,2814	0,1447	-1,94	0,0521
ANO_MODELO2018-2021	0,3807	0,1434	2,66	0,0080
SEXOM	-0,1800	0,0992	-1,81	0,0699

Tabela C.4 – Resultado do ajuste do modelo Tweedie para outras coberturas

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1,6915	0,2141	7,90	0,0000
COD_TARIF2-Passeio Importado	3,6197	0,1390	26,05	0,0000
COD_TARIF3-Pick-up	2,0115	0,1420	14,17	0,0000
COD_TARIF4-Moto	2,6299	0,1491	17,64	0,0000
REGIAOMG-Int	-0,8805	0,1835	-4,80	0,0000
REGIAOMG-RM	-1,2771	0,1857	-6,88	0,0000
REGIAORJ-Int	-0,0815	0,1848	-0,44	0,6593
REGIAORJ-RM	-0,9465	0,1825	-5,19	0,0000
REGIAOSP-int	-2,4300	0,1844	-13,18	0,0000
REGIAOSP-RM	-2,8345	0,1882	-15,06	0,0000
IDADE26 a 35 anos	-1,6951	0,1582	-10,71	0,0000
IDADE36 a 45 anos	-2,3346	0,1595	-14,63	0,0000
IDADE46 a 55 anos	-2,3086	0,1615	-14,30	0,0000
IDADE56 anos e +	-2,5675	0,1623	-15,82	0,0000
ANO_MODELO2011-2013	-0,5649	0,1435	-3,94	0,0001
ANO_MODELO2014-2017	-1,0471	0,1436	-7,29	0,0000
ANO_MODELO2018-2021	-0,3741	0,1415	-2,64	0,0083
SEXOM	-0,3428	0,1005	-3,41	0,0007