



Universidade Federal do Rio de Janeiro (UFRJ)
Instituto de Matemática (IM)
Departamento de Métodos Estatísticos (DME)



Ricardo Boeira Ivanov

Protótipo para precificação de seguros de automóvel a partir de dados públicos utilizando modelos de credibilidade bayesianos de Bühlmann-Straub

MADX02 - Projeto Final em Ciências Atuariais

Rio de Janeiro, 2022

Ricardo Boeira Ivanov

Protótipo para precificação de seguros de automóvel a partir de dados públicos utilizando modelos de credibilidade bayesianos de Bühlmann-Straub

Trabalho apresentado ao Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do Grau de Atuário e de Estatístico.

Universidade Federal do Rio de Janeiro (UFRJ)

Orientadora: Prof.^a Viviana das Graças Ribeiro Lobo, Dr.Sc.

Rio de Janeiro

2022

CIP - Catalogação na Publicação

I93p Ivanov, Ricardo Boeira
Protótipo para precificação de seguros de automóvel a partir de dados públicos utilizando modelos de credibilidade bayesianos de Bühlmann-Straub / Ricardo Boeira Ivanov. -- Rio de Janeiro, 2022.
77 f.

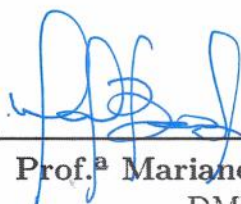
Orientadora: Viviana das Graças Ribeiro Lobo.
Trabalho de conclusão de curso (graduação) - Universidade Federal do Rio de Janeiro, Instituto de Matemática, Bacharel em Ciências atuariais, 2022.

1. Seguro. 2. Autoseg. 3. Credibilidade. 4. Bühlmann-Straub. I. Lobo, Viviana das Graças Ribeiro, orient. II. Título.

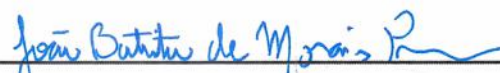
Projeto final de curso defendido e aprovado em 09 de setembro de 2022 pela comissão avaliadora constituída pelos professores e professoras:



**Prof.^a Viviana das Graças Ribeiro
Lobo, Dr.Sc.**
Orientadora, DME/IM/UFRJ



Prof.^a Mariane Branco Alves, Dr.Sc.
DME/IM/UFRJ



**Prof. João Batista de Moraes Pereira,
Dr.Sc.**
DME/IM/UFRJ

Rio de Janeiro, 2022

*Dedico este trabalho a todos aqueles que têm fome e sede de justiça e que veem na
Matemática a forma mais perfeita de exercê-la.*

Agradecimentos

Gratidão. *Do latim, gratitudo. Sentimento de lembrança e agradecimento por um bem recebido, em relação ao autor. Reconhecimento. Qualidade do que é grato.*

Sentimento de lembrança e agradecimento... de quem eu vou me lembrar no futuro quando ler este trabalho e olhar para o passado? Por quem eu nutrirei, de forma atemporal, o sentimento de gratidão por tudo que vivi nessa jornada acadêmica?

Em primeiro lugar, eu vou me lembrar de uma pessoa que eu sequer conheci: o estatístico Kaiser Fung. Foi por conta de um livro dele, comprado de forma bastante despreziosa num aeroporto, enquanto aguardava um voo, que eu descobri o que era Ciências Atuariais e como a Estatística é poderosa. Mesmo já sendo engenheiro químico formado, nunca tinha me atinado à utilidade e importância da Estatística. Sr. Kaiser, muito obrigado pelo seu livro! Foi a partir dele que tudo isso começou. Talvez tenha sido mais uma obra do “*Grande Aleatório*”, que subjaz e a tudo governa no Universo, que fez seu livro chegar às minhas mãos e exercer um impacto tão poderoso, a ponto de me motivar a fazer uma nova graduação.

Depois, vou me lembrar da minha esposa, Robercia, e dos meus pais, Carlos e Maria Isabel. Quando contei essa ideia *maluca* de fazer uma nova graduação, eu recebi muito apoio para iniciar esta jornada. (Sim, houve pessoas que me chamaram de *louco* – mas não foram eles.) Principalmente à Robercia, com quem vivo todos os dias, e ao meu filho Bernardo, que tinha pouco mais de 1 ano quando iniciei os estudos e agora já tem 7 anos e consegue ler este agradecimento, agradeço muito pelo amor e pela paciência. Não foram fáceis esses 6 anos em que eu saía logo cedo de casa e voltava tarde da noite. Muitos foram os dias que ao sair, o Bernardo ainda dormia; e ao retornar, já adormecera. Mas agora essa saga chegou ao fim. Muito obrigado e amo vocês!

Aos professores do Instituto de Matemática meu eterno muito obrigado. Aprendi muito com todos vocês! Alguns nomes eu faço questão de deixar registrado, pois vocês marcaram muito a minha vida ou a minha formação acadêmica: Nei, Mariane, João Ismael, João Batista, Rafael, Carlos Renato, Ralph e Kelly. Vou me lembrar de cada um de vocês com muita gratidão. Talvez vocês não saibam exatamente como e nem porque, mas uma frase dita; um novo olhar para o mundo que vocês me passaram; um choque de realidade que vocês me impuseram; um desafio; ou simplesmente o modo de ser de vocês exerceu um poderoso impacto em mim. Depois que os conheci, não fui mais o mesmo. Muito obrigado!

Certamente vou me lembrar com muito carinho da Prof.^a Viviana. Ela teve o privilégio de finalizar a minha jornada acadêmica: ministrou a minha última disciplina e também aceitou orientar este trabalho de conclusão de curso. Muito obrigado pelos

ensinamentos, pela paciência e pela motivação nesta reta final!

Althayr Santos e Vinicius Grijó, amigos para a vida toda! Que improvável esta amizade, né? Tantas características distintas, como que pôde ter dado certo? Como que essa amizade iniciou, quando todas as circunstâncias não eram favoráveis para este evento? Ah, o aleatório... como é poderoso o aleatório! Ele se encarregou que nos conhecêssemos, mas ele não poderia edificar uma amizade por conta própria. Ele só lançou a semente... mas para a semente germinar, precisa cair em terra boa. Vou carregar nossa amizade pelo resto da vida e vou me lembrar de nossas eternas conversas e risadas, entre um café e outro, enquanto aguardávamos mais uma aula. Muito obrigado pela amizade de vocês!

Por fim, agradeço a todos os funcionários efetivos e terceirizados da UFRJ, em seus diversos setores: limpeza, vigilância, jardinagem, manutenção, alimentação, biblioteca e administrativo, os quais se empenham diariamente para manter a universidade em condições adequadas para que todos possam estudar, mesmo diante de tantas dificuldades que a realidade de nosso país a impõe.

“Não andeis, pois, inquietos, nem digais: que havemos de comer? Que havemos de beber? Com que nos havemos de vestir? Os mundanos é que se preocupam com todas estas coisas. Vosso Pai Celeste sabe que de tudo isto haveis mister. Buscai, pois, em primeiro lugar o reino de Deus e sua justiça, e todas estas coisas vos serão dadas de acréscimo. Não andeis, portanto, solícitos pelo dia de amanhã; o dia de amanhã cuidará de si mesmo; basta a cada dia a sua lida”. (Jesus Cristo)

A estatística sabe de tudo.

É considerada precisamente a quantidade de terra arável na URSS com uma subdivisão em terras negras, siltosas e floresta. Todos os cidadãos, de ambos os sexos, são registrados em livros grossos, tão conhecidos por Ippolit Matveevich Vorobyandinov – livros de cartórios. Sabe-se quanta comida come por ano um cidadão médio da república. Sabe-se o quanto de vodca este cidadão bebe em média com uma indicação aproximada da refeição consumida. Sabe-se quantos caçadores, bailarinas, tornos mecânicos de cabrestante, cães de todas as raças, bicicletas, monumentos, garotas, faróis e máquinas de costura estão no país.

Quanta vida, cheia de ardor, paixões e pensamentos, está nos olhando das tabelas estatísticas!

(...)

Das estatísticas não é possível se esconder em nenhum lugar. Ela tem informações precisas não apenas sobre o número de dentistas, máquina de enchimento de salsicha, zeladores, cineastas, prostitutas, telhados de palha, viúvas, cocheiros e sinos – mas ela sabe até mesmo quantos estatísticos existem no país.

(...)

Ostap não ficou perturbado com o fracasso desta quarta cadeira. Ele conhecia todas as pequenas coisas do destino.

“A felicidade – argumentou ele – sempre chega no último minuto. Se você no mercado de Smolensk precisar pegar o bonde número 4 e lá, exceto pelo quarto, passam ainda o quinto, décimo sétimo, décimo quinto, trigésimo, trigésimo primeiro, B, G e duas linhas de ônibus, então esteja certo que passará primeiro o G, depois dois décimos quintos seguidos, o que geralmente é antinatural, depois o décimo sétimo, o trigésimo, muitos B, novamente G, trigésimo primeiro, quinto, novamente décimo sétimo e novamente B. E agora, quando você começa a pensar que o quarto número não existe mais na natureza, ele chegará lentamente dos lados da estação de Bryansk, coberto com pessoas. Mas para um passageiro de bonde qualificado entrar no vagão não é difícil. Só é necessário que o bonde chegue. Se você precisar entrar no décimo quinto, não duvide: virão primeiramente muitos carros de todos os outros números, o maldito quarto passará oito vezes seguidas e o décimo quinto, que até recentemente vinha a cada cinco minutos, não aparecerá mais que uma vez por dia. É preciso apenas paciência e você vai esperar”.

(Extraído de *A Aventura das 12 Cadeiras*, novela clássica russa escrita pelos autores soviéticos de Odessa, Ilya Ilf e Evgueni Petrov. Tradução direta do original russo por Reinaldo Guarino, Editora Pomnite, São Paulo, 2021.)

Resumo

O seguro de automóvel é o principal produto do mercado segurador brasileiro, tendo alcançado 29% de todos os prêmios emitidos no ano de 2021 segundo dados da Superintendência de Seguros Privados (Susep). Atualmente, quando um cidadão deseja contratar um seguro, ele dispõe de simuladores de preço que quase sempre estão vinculados a corretoras ou seguradoras, necessitando prestar uma série de informações pessoais para ter acesso a estimativas produzidas pelas partes interessadas em vender seguros. A partir de dados públicos do Autoseg da Susep, este trabalho tem o objetivo de construir um protótipo de aplicativo no `RStudio Shiny` para precificação de seguros, disponibilizando à sociedade uma ferramenta independente das seguradoras. Para o estudo, foram selecionados 17 modelos de veículos no horizonte temporal de 2012 a 2018, cuja modelagem estatística dos prêmios de risco de frequência-severidade foi realizada segundo modelos de credibilidade de Bühlmann-Straub com abordagem bayesiana. Os resultados das previsões do modelo para o ano de 2019 foram considerados satisfatórios, atingindo-se uma sinistralidade média global de 0,88, havendo sucesso na construção e disponibilização do protótipo na Internet.

Palavras-chaves: Seguro. Autoseg. Credibilidade. Bühlmann-Straub.

Abstract

Auto insurance is the main product of the Brazilian insurance market, having reached 29% of all premiums issued in 2021 according to Superintendence of Private Insurance (Susep). Currently, when a citizen wants to take out insurance, he has price simulators that are almost always linked to brokers or insurance companies, needing to provide a series of personal information to have access to estimates produced by interested parties in sell insurance. Based on public data from Susep's Autoseg, this work aims to build an application prototype in **RStudio Shiny** for insurance pricing, providing society with a tool independent of insurance companies. For the study, 17 vehicle models were selected in the time horizon from 2012 to 2018, whose statistical modeling of claim frequency and severity risk premiums was performed according to Bühlmann-Straub credibility models with a Bayesian approach. The results of the model forecasts for the year 2019 were considered satisfactory, reaching a global average loss ratio of 0.88, having success in the deploy of the prototype on the Internet.

Key-words: Insurance. Autoseg. Credibility. Bühlmann-Straub.

Lista de ilustrações

Figura 1	– Visualização do módulo de análise exploratória do <i>app</i> Autocheck. . . .	47
Figura 2	– Visualização do <i>app</i> Autocheck para precificação de seguros.	52
Figura 3	– <i>Boxplot</i> da frequência de colisão parcial ou perda total do <i>Volkswagen Gol</i> ponderada por unidades expostas em função do sexo, faixa etária do condutor e subgrupo do modelo (seleção Brasil e todos os anos de modelo).	66
Figura 4	– <i>Boxplot</i> das indenizações de roubo ou furto do <i>Chevrolet Corsa</i> ponderada por unidades expostas em função do sexo, faixa etária do condutor e subgrupo do modelo (seleção Brasil e todos os anos de modelo). . . .	67
Figura 5	– Mapa do estado de São Paulo da frequência de outras coberturas do <i>Volkswagen Gol</i> ponderada por unidades expostas em função do subgrupo do modelo e município.	67
Figura 6	– Distribuição das indenizações de colisão com perda total do <i>Fiat Palio</i> ponderada por unidades expostas em função do subgrupo do modelo e grupo de idade do veículo (seleção Brasil e todos os anos de modelo). .	68

Lista de tabelas

Tabela 1	– Modelos de veículos novos e usados mais vendidos no ano de 2018, contemplando 17 modelos distintos selecionados para o protótipo (UOL, 2019) e (AUTOO, 2019).	25
Tabela 2	– Classificação dos municípios em grupos de risco segundo valor da razão de frequência/expostos para cada cobertura.	29
Tabela 3	– Classificação das regiões em grupos de risco segundo valor da razão de frequência/expostos para cada cobertura.	31
Tabela 4	– Parâmetros a_0 e b_0 da distribuição <i>a priori</i> para $\mu_0(\theta_i)$.	43
Tabela 5	– Parâmetros a_0 e b_0 da distribuição <i>a priori</i> para λ_0 dos veículos <i>Fiat Argo, Jeep Compass e Renault Kwid</i> .	45
Tabela 6	– Variáveis identificadas na análise exploratória para modelagem da frequência e severidade.	50
Tabela 7	– Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, sintetizados por modelo de veículo.	54
Tabela 8	– Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, sintetizados por cobertura.	54
Tabela 9	– Desempenho das previsões para a frequência $\lambda(\theta_i)$ considerando intervalos de credibilidade bilaterais com 95% de probabilidade.	55
Tabela 10	– Desempenho das previsões para a severidade $\mu_t(\theta_i)$ considerando intervalos de credibilidade bilaterais com 95% de probabilidade.	55
Tabela 11	– Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, detalhados por cobertura.	71
Tabela 12	– Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, detalhados por modelo de veículo.	74

Lista de abreviaturas e siglas

Autoseg	Sistema de Estatísticas de Automóveis da Susep
CEP	Código de Endereço Postal
Contran	Conselho Nacional de Trânsito
Dpvat	Danos pessoais causados por veículos automotores de via terrestre
Fipe	Fundação Instituto de Pesquisas Econômicas
IBGE	Instituto Brasileiro de Geografia e Estatística
MCMC	Monte Carlo via Cadeias de Markov
Minfra	Ministério da Infraestrutura
SES	Sistema de Estatísticas da Susep
Susep	Superintendência de Seguros Privados

Lista de símbolos

θ	Um parâmetro qualquer (geralmente, indexador de algum modelo de probabilidade)
$\xi(\theta)$	Função densidade de probabilidade <i>a priori</i> para um parâmetro θ
$\xi(\theta \tilde{x})$	Função densidade de probabilidade <i>a posteriori</i> para um parâmetro θ
$l(\theta; \tilde{x})$	Função de verossimilhança
P^{cred}	Prêmio de credibilidade
P^{ind}	Prêmio individual
P^{col}	Prêmio coletivo
ω	Fator de credibilidade
S_{it}	Sinistro para o grupo de risco i no tempo t
v_{it}	Volume de unidades expostas a risco para o grupo de risco i no tempo t
X_{it}	Sinistro ponderado por unidade de volume para o grupo de risco i no tempo t
N_{it}	Frequência de sinistros para o grupo de risco i no tempo t
F_{it}	Frequência de sinistros ponderada por unidade de volume para o grupo de risco i no tempo t
μ	Nível médio de severidade (parâmetro das distribuições normal e log-normal)
β	Fator de crescimento do modelo de severidade
σ^2	Parâmetro associado à variância dentro de um grupo de risco
τ^2	Parâmetro associado à variância entre os grupos de risco e ao fator de crescimento
λ	Taxa de ocorrência de sinistros (parâmetro da distribuição Poisson)
$\mathbb{E}(\cdot)$	Esperança matemática
$Var(\cdot)$	Variância

iid	Independente e identicamente distribuído
$\mathcal{N}(\cdot, \cdot)$	Modelo de probabilidade normal
$\mathcal{LN}(\cdot, \cdot)$	Modelo de probabilidade lognormal
$\mathcal{U}(\cdot, \cdot)$	Modelo de probabilidade uniforme
$\Gamma(\cdot, \cdot)$	Modelo de probabilidade gama
$Poi(\cdot)$	Modelo de probabilidade Poisson

Sumário

1	INTRODUÇÃO	18
2	MATERIAIS E MÉTODOS	23
2.1	RStudio Shiny	23
2.2	Base de Dados Autoseg da Susep	23
2.2.1	Sobre o Autoseg	23
2.2.2	Extração e Filtragem dos Dados	24
2.2.3	Detecção e Tratamento de Inconsistências nos Dados	25
2.2.3.1	Dados faltantes	25
2.2.3.2	Inconsistências de Anos de Modelos	26
2.2.3.3	Problemas nas Informações de Município e CEP	26
2.2.4	Coberturas disponibilizadas no Autoseg	27
2.2.5	Exposição a Risco	28
2.3	Análise exploratória e seus desdobramentos	28
2.3.1	Agrupamento de Municípios	29
2.3.2	Agrupamento de Idade de Veículo	30
2.3.3	Agrupamento de Regiões: vinculação entre as tabelas de Sexo e Faixa Etária com Município	30
2.3.4	Desagrupamento das Coberturas de Colisão Parcial e Colisão Perda Total	31
2.3.5	Agrupamento das Coberturas Roubo/Furto e Incêndio	31
2.4	Modelagem Estatística	32
2.4.1	Estatística Bayesiana	33
2.5	Teoria da Credibilidade	35
2.5.1	Modelo de Bühlmann-Straub	38
2.6	Modelagem para o Protótipo	39
2.6.1	Modelo para Severidade	40
2.6.2	Modelo para Frequência	44
2.6.3	Modelo para Frequência-Severidade	45
3	RESULTADOS E DISCUSSÃO	47
3.1	Módulo Análise Exploratória	47
3.1.1	Cobertura Roubo ou Furto	48
3.1.2	Cobertura Colisão Parcial ou Colisão Perda Total	48
3.1.3	Cobertura Incêndio	49
3.1.4	Outras Coberturas	49
3.1.5	Resumo das Constatações	50

3.1.6	Codificação dos Grupos de Risco	51
3.2	Módulo Precificação	52
3.2.1	Principais Resultados	52
3.2.2	Avaliação do Modelo	53
3.3	Discussão Global dos Resultados	54
3.3.1	Modelo para Frequência de Sinistros	55
3.3.2	Modelagem de Veículos Novos	56
3.3.3	Fenômeno de Nascimento, Migração e Morte de Grupos de Risco	57
3.3.4	Covariáveis ausentes e não utilizadas	58
4	CONCLUSÕES	60
	Sugestões para Trabalhos Futuros	61
	Referências	62
	APÊNDICES	65
	APÊNDICE A – EXEMPLOS DE GRÁFICOS DA ANÁLISE EX- PLORATÓRIA	66
	APÊNDICE B – CODIFICAÇÃO DOS SUBGRUPOS DOS MO- DELOS DE VEÍCULOS	69
	APÊNDICE C – RESULTADOS DE PRÊMIOS E INDENIZAÇÕES PARA O ANO DE 2019 POR COBERTURA	71
	APÊNDICE D – RESULTADOS DE PRÊMIOS E INDENIZAÇÕES PARA O ANO DE 2019 POR MODELO DE VEÍ- CULO	74

1 Introdução

O seguro de automóvel¹, em conjunto com outros seguros diretamente associados a ele², é o maior produto do mercado segurador brasileiro. Segundo dados do [Sistema de Estatísticas](#) (SES) da Superintendência Nacional de Seguros Privados (Susep), as seguradoras emitiram, no ano de 2021, R\$ 38,5 bilhões em prêmios de seguro associados ao ramo auto³, representando 28,97% de todos os prêmios emitidos naquele ano no mercado segurador brasileiro – números que superam o ramo vida, por exemplo, onde foram emitidos R\$ 17,6 bilhões em prêmios ([SUSEP, 2021](#)).

Além de sua importância devida ao alto volume de prêmios emitidos, o ramo auto notabiliza-se por ser um seguro de alta capilaridade nacional: em 2019, 14.093.534 veículos segurados pertenciam a pessoas físicas, enquanto 2.125.536 veículos a pessoas jurídicas⁴.

Ao levar em consideração os dados da frota nacional de veículos automotores, constata-se que a proporção de veículos segurados ainda é baixa. Segundo dados do Ministério da Infraestrutura, no mês de março de 2020 o Brasil tinha uma frota de 105.625.858 veículos registrados, composta por cerca de 60% de veículos leves (automóveis, camionetas e utilitários) ([MINFRA, 2020](#)). Tomando-se os números globais, em torno de apenas 16% dos veículos brasileiros possuem seguro.

Quando comparado a outros países, como os Estados Unidos, conclui-se que ainda há muitas oportunidades a serem exploradas neste ramo. Nos Estados Unidos, em 2019, havia cerca de 276 milhões de veículos automotores registrados, sendo que quase todos possuem seguro⁵. Isso faz com que os prêmios emitidos em seguro auto, nos Estados Unidos, alcancem cifras muito maiores que as do Brasil: em 2019, foram emitidos US\$ 239 bilhões em prêmios de seguro auto naquele país ([POLICYGENIUS, 2022](#)), contra os já mencionados R\$ 38,5 bilhões em prêmios do Brasil, em 2021.

A baixa proporção de veículos segurados na frota nacional pode ter as mais diversas causas, tais como o baixo poder aquisitivo da população, alta tarifação dos seguros no Brasil, não exigência de seguro auto obrigatório e até a idade da frota de veículos

¹ Neste trabalho, o termo “automóvel” (ou simplesmente “auto”, de forma sucinta) será empregado de forma genérica para designar a cobertura de casco dos seguros de veículos automotores, o que inclui automóveis, camionetas, caminhonetes, motocicletas, caminhões, ônibus e utilitários.

² Como de responsabilidade civil ou específico para vítimas de acidentes de trânsito.

³ Inclui os ramos de cód. 0520 a 0589 do SES da Susep.

⁴ Segundo dados do Autoseg da Susep, há ainda outros 164.376 veículos sem informação sobre a titularidade ([SUSEP, 2020](#)).

⁵ O seguro auto é obrigatório, por lei, em 49 dos 50 estados que compõem os Estados Unidos (somente em *New Hampshire* o seguro não é obrigatório por lei, desde que o proprietário demonstre capacidade financeira) ([STATISTA, 2022](#)). No Brasil, o único seguro de natureza obrigatória associado ao ramo auto é o de danos pessoais causados por veículos automotores de via terrestre (Dpvt) – destinado a indenizar as vítimas de acidentes de trânsito.

automotores. Sobre este último aspecto, o Ministério da Infraestrutura aponta que, em outubro de 2021, praticamente a *metade* dos veículos registrados foram fabricados até o ano de 2008 (ou seja, tinham 13 anos de idade ou mais); somente 15% dos veículos em circulação tinham sido fabricados de 2017 a 2021 (até 4 anos de idade) e meros 2,4% dos veículos eram *do ano* (MINFRA, 2021). O fato de a frota nacional ser considerada madura, ou velha, é um dos fatores que reduzem drasticamente a proporção de veículos segurados no Brasil, pois “o carro vai ficando velho e depreciado, porém o valor das peças continua muito parecido com o de veículos novos ou seminovos”, segundo as palavras de Fábio Leme, vice-presidente técnico da HDI Seguros em 2018 (UOL, 2018), o que encarece o seguro e torna o produto não atrativo aos proprietários.

Neste cenário, e concorrentemente aos seguros, existe no Brasil o mercado da *proteção veicular*, um produto ofertado por cooperativas ou associações com a tendência de ser mais barato que um seguro. Conceitualmente, a proteção veicular não é um seguro e não está sujeita à regulamentação da Susep, mas, na prática, emula o funcionamento de um seguro: o proprietário paga uma contribuição mensal ou anual (análoga ao prêmio) e o grupo de associados o indeniza caso ocorra alguma perda aleatória, como uma colisão, roubo etc. Por não ser um seguro, a proteção veicular não traz as garantias que um seguro possui – mas por ser um produto mais barato e também por abarcar carros antigos, acabou popularizando-se no Brasil (AUTOPAPO, 2021).

Dada a relevância do seguro auto, uma série de trabalhos são desenvolvidos sobre a tarifação deste tipo de produto ou sobre a modelagem dos sinistros: análise espacial de prêmios de seguros (TEIXEIRA; SCALON, 2014), previsão de sinistros (ZANIBONI; MONTINI, 2015), tarifação de seguros em região específica do país (FARIAS; JESUS, 2019), identificação de variáveis que melhor discriminam a ocorrência de sinistros (CESAR, 2020), influência do modelo do veículo na precificação de seguros (AZEVEDO, 2020), identificação de grupos de risco para roubos, furtos, perda total ou parcial (PRAZERES, 2020), dentre outros. Quase que a totalidade desses estudos são de interesse maior às seguradoras, por apresentarem modelos matemáticos com o intuito de melhorar a precificação de prêmios ou então a explicação (ou previsão) da ocorrência de sinistros⁶.

De outro lado, os segurados não dispõem de ferramentas que expliquem os preços dos seguros ou então que sirvam como referencial, ou parâmetro, para que eles avaliem os preços informados pela seguradora⁷. Neste sentido, o cálculo dos prêmios de seguro é pouco ou nada transparente para o cidadão comum. Embora muitos segurados entendam quais fatores podem afetar o cálculo do prêmio (como idade, modelo do veículo, município

⁶ Além disso, é importante mencionar que quase todos esses estudos foram realizados com os dados da base Autoseg da Susep.

⁷ Ou seja, não há uma informação fornecida por uma terceira parte, em referenciais minimamente comparativos, para que o segurado possa avaliar se os preços praticados pelas seguradoras fazem certo sentido ou não.

de circulação etc.), poucos entendem o porquê dos valores apresentados a título de prêmio.

É bastante comum que as corretoras de seguro e as seguradoras disponibilizem ferramentas para o segurado simular os prêmios de seguro auto. Inevitavelmente, essas ferramentas trazem consigo os modelos matemáticos que as próprias seguradoras empregam; ou seja, em essência, o segurado somente consegue efetuar comparações de preços entre as diversas seguradoras, mas não possui uma informação externa, fornecida por uma terceira parte sem qualquer interesse na contratação do seguro, sobre a estimativa do prêmio de seguro, conforme seu perfil.

Soma-se a isso o fato que essas ferramentas de simulação de preços de seguro fatalmente solicitam informações pessoais do interessado em algum momento da simulação, sem muita clareza sobre como elas são tratadas ou utilizadas. Ou seja, não fica claro se as informações efetivamente são levadas em conta para simular um preço ou se elas servem tão-somente para coleta de dados do interessado, a fim de que este seja contatado, futuramente, por um corretor de seguros que tentará vender uma apólice. Um exemplo de simulador (*Smartia Seguros Online*) solicita a data de nascimento, nome completo e e-mail da pessoa que faz a simulação; a empresa *Minuto Seguros* solicita a placa do veículo para fazer uma cotação de seguro, sob a alegação de que “*com a placa, conseguimos conhecer mais sobre seu carro*”; os simuladores das seguradoras vinculadas a bancos, muitas das vezes integrados ao *Internet banking*, já dispõem de muitos dados dos clientes; o simulador da empresa *Youse* solicita, como primeira informação para utilizar o simulador, o nome e e-mail do interessado, informando que “*a gente usa estas informações pra ativar, simular ou cotar os produtos Youse que você tem interesse. Nestes casos, também podemos usar os seus dados pra esclarecer dúvidas e até mesmo oferecer oportunidades que combinem com o seu perfil*”; e, por fim, a plataforma *Cotação Seguro*, a qual faz cotações em diversas seguradoras, exige que o contato seja realizado por *Whatsapp*.

Em suma, além de não haver⁸ uma forma de cotar seguros auto desvinculada de interesses comerciais, os interessados ainda necessitam informar seus dados pessoais para terem acesso a estimativas de prêmios de seguro. Nasce, assim, a motivação e relevância do presente trabalho.

A partir da base de dados pública *Autoseg* da Superintendência de Seguros Privados (SUSEP, 2020), será elaborado e apresentado um protótipo para precificação de seguros auto (cobertura casco), implementado na biblioteca *Shiny* do *software R* (R CORE TEAM, 2021). Por se tratar de um protótipo, somente alguns modelos de veículos foram selecionados para fazer parte do estudo, tendo sido selecionados os carros novos e usados mais vendidos e emplacados no ano de 2018 (UOL, 2019) e (AUTOO, 2019). Para esses modelos de veículos, adotando-se uma modelagem estatística bayesiana, o protótipo precificará os seguros para as coberturas contra roubo, furto, incêndio, colisão parcial, colisão

⁸ Ou ser difícil de encontrar.

com perda total e outras coberturas (como assistência 24 horas, blindagem, vidros etc.), levando em conta algumas informações do perfil do condutor e do veículo.

Para estimar o prêmio de um seguro, é fundamental conhecer a *frequência* de ocorrência dos sinistros e sua *severidade*, onde a combinação entre essas variáveis possibilita o cálculo do prêmio de risco de uma apólice. O cerne do problema de tarifação envolve fazer previsões sobre a frequência e a severidade de forma satisfatória, a partir de algumas características conhecidas do segurado: sexo, idade, modelo do veículo, se o segurado possui garagem na residência, onde ele mora, se utiliza o carro para trabalhar etc.⁹ Obviamente, como neste trabalho os dados são provenientes de uma base de dados pública – onde não estão disponíveis todas as informações sobre o segurado –, o desafio reside, também, em ajustar modelos de precificação razoáveis a partir de informações em níveis agregados e com menos variáveis disponíveis.

No escopo deste trabalho, os dados extraídos do Autoseg compreenderão os anos de 2012-2019, sendo que serão utilizados os anos de 2012 a 2018 para fins de modelagem estatística dos prêmios, reservando-se o ano de 2019 para testar as estimativas produzidas pelo modelo¹⁰.

A construção do protótipo envolve uma série de etapas de análise de dados, partindo desde a extração dos dados da base Autoseg, filtragem e agrupamento dos dados de estudo, passando pela realização de uma análise exploratória, a fim de colher *insights* para a seleção de modelos estatísticos, até culminar no desenvolvimento da ferramenta para consulta dos prêmios.

Neste documento, consta a descrição de todas as premissas e hipóteses adotadas para essas etapas. Em função do alto volume de dados, e também da quantidade de veículos em estudo, não serão apresentados gráficos e tabelas. O leitor pode consultar os gráficos e resultados completos deste trabalho acessando os dois aplicativos **Shiny** desenvolvidos: [Autocheck - Módulo Análise Exploratória](#)¹¹ e [Autocheck - Precificação](#)¹².

O presente trabalho está organizado da seguinte forma: o Capítulo 2 apresenta uma descrição dos métodos empregados para analisar os dados, incluindo uma revisão

⁹ Algumas características do segurado não são propriamente conhecidas pela seguradora, mas sim obtidas a partir de uma autodeclaração do segurado; também, outras características que podem vir a ser relevantes sequer são coletadas pela seguradora. Naturalmente, a relação entre seguradora e segurado é sujeita à assimetria de informação, um aspecto relevante na tarifação de seguros e que não será abordado neste estudo.

¹⁰ Até junho de 2022, os últimos dados disponibilizados no Autoseg ainda eram do 1º semestre de 2020, fazendo com que 2019 seja o último ano com dados anuais fechados. Em pedido de informação protocolado à Susep por este autor, em 25 de novembro de 2021 (nº 18830.001981/2021-38), a superintendência informou que “em função da pandemia de Covid-19, foram postergados os prazos para envio dos dados referentes ao seguro de automóveis” e que “foram realizadas modificações nos sistemas de recepção, crítica, armazenamento e disponibilização de dados, o que também ocasionou atraso na publicação dos dados de Autoseg”.

¹¹ <https://rbivanov.shinyapps.io/autocheck-TCC-AED/>

¹² <https://rbivanov.shinyapps.io/autocheck-tcc>

bibliográfica sucinta abordando a modelagem estatística bayesiana e a Teoria da Credibilidade, seguindo do enunciado da modelagem estatística empregada para o estudo. Já no Capítulo 3, constam os resultados e uma discussão sobre as estimativas produzidas pelo modelo. Por fim, no Capítulo 4 constam as conclusões deste estudo, seguidas por sugestões para trabalhos futuros.

2 Materiais e Métodos

2.1 RStudio Shiny

O Shiny é um produto da RStudio que permite criar aplicativos¹ interativos para a *web* diretamente do R. Os aplicativos podem ser hospedados de forma independente em páginas na Internet ou incorporados a documentos R Markdown em `html`.

Uma forma bastante prática para hospedar aplicativos Shiny é por intermédio do `shinyapps.io` do RStudio, cujo uso é bastante simples. O desenvolvedor pode fazer todo o processo de disponibilização do aplicativo dentro do próprio RStudio mediante comandos, fazendo a configuração do servidor no *site* do [ShinyApps](#).

O `shinyapps.io` possui diversos planos pagos para os desenvolvedores, sendo que quanto mais cara é a licença, mais recursos ficam disponíveis ou são maiores, desde horas de máquina (consumidas quando o *app* está em execução) até suporte ou autenticação para uso do *app*. As licenças pagas variam de US\$ 9,00 a US\$ 299,00 por mês².

Para a hospedagem dos *apps* do presente trabalho, foi utilizada a licença gratuita do `shinyapps.io`, a qual permite a hospedagem de até 5 *apps* com 25 horas disponíveis para uso, por mês. Na licença gratuita, quando o limite de horas é alcançado, o *app* fica indisponível para uso, devendo-se aguardar o próximo mês³.

2.2 Base de Dados Autoseg da Susep

2.2.1 Sobre o Autoseg

A Susep, criada no ano de 1966 e atualmente vinculada ao Ministério da Economia, é uma autarquia responsável pelo controle e fiscalização dos mercados de seguro, previdência privada aberta, capitalização e resseguro.

Desde o ano de 2006, a partir de arquivos enviados semestralmente pelas companhias seguradoras, a Susep disponibiliza dados estatísticos do seguro de automóveis, referentes a apólices vigentes e sinistros ocorridos⁴. Além de uma ferramenta para con-

¹ Neste trabalho, será usado o termo *app* para designar esses aplicativos.

² Valores consultados em 25/06/2022 para as licenças *starter* e *professional*.

³ O limite de 25 horas é para uma conta e computa conjuntamente o tempo consumido em todos os *apps* desenvolvidos quando acessados via Internet.

⁴ Segundo a Susep, “até a versão disponibilizada em junho/2005, o sistema era fornecido em CD-ROM, de forma que o usuário precisava instalar o programa e a base de dados completa em seu computador. Os procedimentos necessários à instalação se revelaram demasiadamente complexos para o usuário comum” (SUSEP, 2020).

sulta online, a Susep também disponibiliza os arquivos de dados para *download* (SUSEP, 2020).

“O sistema fornece informações sobre número de veículos expostos, prêmio médio, importância segurada média, número de sinistros e valor de indenizações, classificadas de acordo com categoria, modelo e ano do veículo, região ou CEP de circulação, e perfil do segurado” (SUSEP, 2020).

A Susep segrega os dados de exposição, prêmios, sinistros e categoria tarifária para os modelos de veículos e cada ano de modelo⁵ em três tabelas, as quais são classificadas pelas seguintes chaves:

- Região, sexo ou faixa etária;
- CEP;
- Município.

Os dados brutos do Autoseg não individualizam cada apólice, mas sim agrupam os dados até certo nível, não sendo possível associar as três tabelas plenamente entre si. Excetuando as tabelas com CEP e município, para as quais é possível estabelecer uma associação, a tabela de região, sexo e faixa etária está dissociada das demais; por exemplo, não é possível saber o sexo ou a faixa etária dos condutores de *Fiat Palio* em Osasco, mas tão-somente para a região metropolitana de São Paulo (que abarca mais de um município)⁶.

2.2.2 Extração e Filtragem dos Dados

Os dados foram baixados manualmente do Autoseg e tratados no R com a biblioteca `tidyverse`. O Autoseg disponibiliza arquivos de dados semestrais, tendo sido extraídos 48 arquivos semestrais (compreendendo o período de 2012-1 até 2019-2), onde cada tabela semestral de dados possui um tamanho médio de 885 MB (por CEP), 340 MB (por município) e 269 MB (por sexo e faixa etária), totalizando 23,6 GB de dados.

Depois de extraídos, os dados foram filtrados selecionando-se somente os modelos de interesse para a construção do protótipo, os quais estão apresentados na Tabela 1. Em seguida, os dados dos diversos códigos Fipe de uma mesma marca/modelo foram agrupados por similaridade. Sobre este último aspecto, cabe uma explanação mais detalhada.

Segundo a Resolução Contran nº 916, de 28 de março de 2022, “*todos os veículos fabricados, montados e encarroçados, nacionais ou importados, devem possuir código de*

⁵ A Susep detalha o veículo segundo sua marca, modelo, versão e ano, usando a mesma codificação da Fipe. Neste trabalho, serão analisados 17 modelos, mas é importante destacar que para cada modelo de veículo há uma série de versões disponíveis.

⁶ A Susep estabeleceu uma divisão do Brasil em 41 regiões, as quais agrupam municípios próximos – não necessariamente contíguos – por aparente similaridade de risco.

Tabela 1 – Modelos de veículos novos e usados mais vendidos no ano de 2018, contemplando 17 modelos distintos selecionados para o protótipo (UOL, 2019) e (AUTOO, 2019).

Veículos Novos	Veículos Usados
1. Chevrolet Onix	1. Volkswagen Gol
2. Hyundai HB20	2. Fiat Uno
3. Ford Ka	3. Fiat Palio
4. Volkswagen Gol	4. Ford Fiesta
5. Chevrolet Prisma	5. Chevrolet Celta
6. Volkswagen Polo	6. Fiat Strada
7. Renault Kwid	7. Chevrolet Corsa
8. Fiat Strada	8. Volkswagen Fox
9. Fiat Argo	9. Fiat Siena
10. Jeep Compass	10. Ford Ka

marca/modelo/versão específico, o qual deve ser concedido conjuntamente à emissão, pelo órgão máximo executivo de trânsito da União, do Certificado de Adequação à Legislação de Trânsito (CAT)”. Ocorre que os fabricantes de veículos registram uma série de códigos para veículos de mesmo modelo que possuem sutis diferenças entre si, como de potência, cilindros, cilindrada, quantidade de portas, quantidade de válvulas etc. Exemplificando, na base de dados Autoseg, para o *Chevrolet Celta*, existem 13 versões para *Celta 1.0* e 6 versões para *Celta acima de 1.0*.

Assim, no exemplo dado, ao invés de tratar separadamente cada uma das 19 versões do *Chevrolet Celta*, os dados foram agrupados por similaridade para a análise exploratória e modelagem estatística, adotando-se a própria classificação de *subgrupos* de modelos que a Susep disponibiliza no Autoseg. No caso do *Chevrolet Celta*, o problema resume-se a dois subgrupos.

2.2.3 Detecção e Tratamento de Inconsistências nos Dados

Antes de iniciar a análise exploratória, os dados passaram por um último tratamento, a fim de detectar possíveis dados faltantes, problemas de formatação ou inconsistências. Diversos problemas foram encontrados, conforme detalhado nas seções a seguir.

2.2.3.1 Dados faltantes

Para as variáveis relacionadas ao sexo do condutor, faixa etária e município, a base Autoseg possui um código específico para dado faltante (*não informado*), cujos casos foram eliminados da análise.

Também, foram detectados anos de veículos preenchidos com 0 ou 1900, os quais também foram interpretados como sendo dados faltantes e, assim, eliminados do estudo.

Por fim, o caso mais grave de dados faltantes ocorreu na tabela segregada por sexo e faixa etária, onde há também a identificação dos veículos de pessoas jurídicas⁷. Ainda que o volume de veículos expostos de pessoas jurídicas seja considerável⁸, os valores de indenização encontram-se zerados em boa parte dos arquivos explorados, para todas as coberturas. Desta maneira, os dados de veículos de pessoas jurídicas tiveram de ser removidos do estudo.

2.2.3.2 Inconsistências de Anos de Modelos

Ao explorar os dados, também foram eliminadas as inconsistências nos dados de anos de modelo dos veículos. Havia dados onde o ano do veículo era maior que o ano do envio dos dados; também, houve casos onde o ano do modelo constava fora do intervalo de fabricação dos modelos por parte das montadoras. Esses dados inconsistentes foram todos removidos do estudo⁹.

2.2.3.3 Problemas nas Informações de Município e CEP

As três tabelas do Autoseg possuem informações sobre a localização da apólice, a saber:

- Tabela por sexo e idade: consta a informação da região, conforme classificação e agrupamento da Susep;
- Tabela por município: consta o município, conforme codificação de municípios da Susep;
- Tabela por CEP: consta os primeiros 5 algarismos do CEP (base dos Correios).

Inicialmente, detectou-se um problema na codificação de municípios do Autoseg: a Susep não adota os mesmos códigos do Instituto Brasileiro de Geografia e Estatística (IBGE). Além dessa questão, a Susep lista 3450 possíveis municípios no Brasil, enquanto as tabelas do IBGE indicam que o Brasil possui 5297 municípios cadastrados, indicando que a base do Autoseg não prevê todos os municípios brasileiros¹⁰. Em alguns casos, detectou-se que a Susep coloca estados inteiros como se fossem um município, como, por exemplo, Amazonas, Paraíba, Rio Grande do Norte e Mato Grosso do Sul.

⁷ Esta informação consta na variável de sexo do condutor.

⁸ No ano de 2019, foram mais de 2.000.000 de veículos de pessoas jurídicas segurados no Brasil.

⁹ Exceção feita ao caso em que o ano do modelo do veículo era somente 1 ano posterior ao ano do envio dos dados. Por exemplo, no mercado de automóveis, é comum que um carro com ano de modelo 2023 seja fabricado e comercializado já no ano 2022.

¹⁰ Alguns casos chamam a atenção no Autoseg, como a ausência dos municípios de Serra, Colatina, Cariacica, Cachoeiro de Itapemirim, Linhares, Vila Velha e Aracruz, todos no estado do Espírito Santo e com mais de 100.000 habitantes, que sequer possuem código de município no Autoseg.

Também, a codificação de municípios do Autoseg não possui informação sobre a qual unidade da federação (UF) o município pertence, impossibilitando saber sua real localização geográfica, principalmente ao se considerar que, segundo o IBGE, há 273 municípios homônimos no Brasil (pouco mais de 5%). Há casos em que essa questão é extremamente relevante: por exemplo, *Cachoeirinha* é o nome de um município do estado do Pernambuco e também do Rio Grande do Sul. Porém, o primeiro trata-se de uma pacata cidade com 20.000 habitantes enquanto o segundo está localizado na região metropolitana de Porto Alegre¹¹. Ademais, a base Autoseg possui uma série de nomes de municípios com ortografia incorreta, necessitando de um tratamento manual dos dados.

Problemas similares foram encontrados na tabela segregada por CEP, com o agravante de que na base de CEP da Susep consta somente os primeiros 5 algarismos, o que no caso de municípios menores recai ao próprio município. Abstraindo desses problemas, e depois de efetuar o máximo tratamento e ajuste possível nos dados de município, estes foram utilizados para a construção de mapas para análise exploratória, depois de agregar dados externos com informações de latitude e longitude¹², não tendo sido utilizado o CEP nas análises.

2.2.4 Coberturas disponibilizadas no Autoseg

O Autoseg separa os dados de frequência de sinistros e indenização nas seguintes classes de cobertura:

- Roubo/furto;
- Colisão parcial;
- Colisão perda total;
- Incêndio;
- Outras coberturas (como assistência 24 h, vidros, blindagem, equipamentos acessórios etc.).

Nos arquivos de dados segregados por CEP e por município, as coberturas de colisão parcial e colisão perda total estão consolidadas em uma única cobertura, sem distinção – diferentemente do que ocorre no arquivo de dados segregado por sexo e faixa etária. Desta maneira, a fim de uniformizar a análise exploratória com os diferentes arquivos de dados, as coberturas de colisão parcial e colisão perda total foram tratadas conjuntamente na análise exploratória inicialmente elaborada.

¹¹ Outro exemplo é *Bom Jesus*, nome de cinco diferentes municípios, nos estados do Piauí, Rio Grande do Norte, Paraíba, Santa Catarina e Rio Grande do Sul.

¹² Nos casos de municípios homônimos, foi eleito como estado verdadeiro aquele onde o município era o mais populoso.

2.2.5 Exposição a Risco

O Autoseg possui uma única variável em que apresenta o número de unidades expostas ao risco. Conforme as definições do Autoseg, “*o conceito de exposição leva em conta o tempo em que cada apólice esteve vigente, dentro da janela de observação, que é o período semestral abrangido em cada atualização do Autoseg. Desta forma, o número de expostos, apurado para um período anual, representa o melhor estimador disponível para a quantidade de veículos segurados*”.

É sabido, contudo, que no mercado de seguros há apólices que compreendem somente determinadas coberturas. Como nos dados do Autoseg não existe diferenciação da quantidade de unidades expostas para as diferentes coberturas, partiu-se da premissa que a exposição é a mesma para todos os riscos¹³.

2.3 Análise exploratória e seus desdobramentos

Considerando as particularidades dos dados e o escopo do protótipo, a análise exploratória foi organizada de forma a obter *insights* sobre quais variáveis explicativas aparentam ser mais relevantes para explicar a frequência e severidade dos sinistros, balizando a modelagem, além de verificar aspectos de simetria e presença de valores discrepantes nas observações de frequência e severidade de sinistros.

Em função das diferentes bases do Autoseg (por sexo e faixa etária, e por município), a análise exploratória inicialmente foi dividida nesses dois grandes grupos. Para carregar os dados do *app*, além de selecionar o modelo de veículo de interesse, é necessário especificar qual fonte de dados deseja-se explorar. O usuário pode selecionar uma faixa de período de anos para explorar os dados, mediante o uso de um filtro. Para fins da formação de entendimento neste trabalho, sempre se adotou o período completo de dados (2012-2018) nas análises.

Uma vez carregados os dados, duas variáveis resposta foram inspecionadas:

- Razão de frequência de sinistros por unidades expostas;
- Razão de indenização de sinistros por frequência de sinistros.

Essas variáveis são necessárias para realizar a precificação do seguro, uma vez que o conhecimento e a combinação da *frequência* de ocorrência dos sinistros e de sua *severidade*, quando um sinistro ocorre, possibilita a estimação do prêmio de risco de uma apólice, para um determinado grupo de risco.

¹³ Análogo ao caso em que um veículo possui uma apólice de seguro contemplando todas as coberturas: de colisão, roubo, furto, incêndio e com outros serviços.

Para a fonte de dados por sexo e faixa etária¹⁴, foram construídos gráficos do tipo *boxplot* para as duas variáveis resposta em estudo. Nesses gráficos, cada dado corresponde à totalização das observações das categorias de sexo e faixa etária em cada ano de envio (ver exemplos das Figuras 3 e 4 no Apêndice A).

Para a fonte de dados por município, optou-se pela construção de mapas geográficos com bolhas para as variáveis resposta em estudo, posicionando-as na localização geográfica de cada município. As variáveis resposta foram totalizadas por todo o período de observação (2012-2018), plotando-se, assim, um único ponto por município nos mapas. Para melhor visualização, o tamanho da bolha é proporcional ao valor da variável resposta (frequência por expostos ou indenizações por frequência), enquanto as cores representam o grupo de risco do município, conforme será dissertado posteriormente (ver exemplo da Figura 5 no Apêndice A).

Todos os gráficos e mapas foram separados para os subgrupos de modelo dentro de um mesmo modelo de veículo a fim de comparar a influência dessa variável, além do detalhamento para cada uma das coberturas de risco em estudo. Por fim, um filtro para a idade do veículo pode ser aplicado em todos os gráficos.

2.3.1 Agrupamento de Municípios

A Susep agrupa 3450 municípios brasileiros em 41 regiões de aparente similaridade de risco, definidas pela superintendência. Ainda assim, essa quantidade de níveis em uma variável qualitativa é considerável para o posterior ajuste de modelos estatísticos.

Desta maneira, dado um modelo de veículo e um subgrupo de modelo, e totalizando-se as observações do período 2012-2018, os municípios foram classificados em 3 grupos de risco¹⁵, conforme o critério constante na Tabela 2.

Tabela 2 – Classificação dos municípios em grupos de risco segundo valor da razão de frequência/expostos para cada cobertura.

Cobertura	Baixo Risco	Médio Risco	Alto Risco
Roubo/Furto	[0; 0,001)	[0,001; 0,010)	[0,010, +∞)
Colisão parcial ou total	[0; 0,05)	[0,05; 0,10)	[0,10, +∞)
Outras coberturas	[0; 0,05)	[0,05; 0,10)	[0,10, +∞)

Sob esse critério, por exemplo, um determinado município pode ser classificado de alto risco para roubo/furto de *Volkswagen Gol acima de 1.0* e ser considerado de baixo

¹⁴ Variáveis qualitativas com dois e cinco níveis, respectivamente.

¹⁵ Exceto para a cobertura de incêndio, onde o município não foi considerado relevante para fins explicativos. Neste caso, somente para construção dos mapas geográficos, os municípios foram classificados nos grupos *com sinistro* e *sem sinistro*.

risco para roubo/furto do *Renault Kwid*, uma vez que essa classificação de risco é atribuída quando fixado um modelo e subgrupo de modelo de veículo.

2.3.2 Agrupamento de Idade de Veículo

Para os modelos de veículos selecionados para o protótipo, há casos onde a idade do veículo pode variar bastante, caso do *Volkswagen Gol*, *Chevrolet Corsa* ou *Fiat Palio*, e outros em que os modelos são novos diante do seu surgimento no mercado (caso do *Jeep Compass*, *Fiat Argo* e *Renault Kwid*). De toda forma, incluir a idade do veículo diretamente na modelagem não parece ser produtivo pois, conforme a análise exploratória realizada, é difícil capturar influências na variável resposta a partir de mudanças sutis de 1 ou 2 anos na idade do veículo.

Desta maneira, decidiu-se agrupar as idades dos veículos da seguinte maneira:

- Grupo A: Idade do veículo $\in \{0, 1, 2, 3\}$ anos;
- Grupo B: Idade do veículo $\in \{4, \dots, 10\}$ anos;
- Grupo C: Idade do veículo $\in \{11, \dots, 40\}$ anos.

A Figura 6 constante no Apêndice A apresenta um exemplo do efeito do subgrupo do modelo e grupo de idade do veículo nas indenizações.

2.3.3 Agrupamento de Regiões: vinculação entre as tabelas de Sexo e Faixa Etária com Município

O Autoseg não possui uma conexão entre as tabelas de dados por sexo e faixa etária do condutor com a tabela de dados de municípios. Assim, caso opte-se em modelar a frequência de colisões utilizando-se o município, sexo e faixa etária do condutor como variáveis explicativas, não é possível vincular ambos arquivos de dados diretamente.

Contudo, ambas tabelas apresentam a informação da *região de circulação* segundo classificação adotada pela Susep¹⁶, sendo possível estabelecer uma conexão indireta entre as tabelas utilizando-se esta variável.

Assim, a *região de circulação* da Susep foi adotada em substituição ao município para fins de modelagem matemática. A Tabela 3 apresenta o critério de classificação das regiões em alto, médio e baixo risco.

O critério de classificação das regiões diferiu do critério dos municípios pois as regiões inevitavelmente agregam municípios de diferentes riscos, com a questão de possuírem um volume de veículos expostos bastante superior ao volume de expostos nos municípios.

¹⁶ A qual divide o Brasil em 41 regiões.

Tabela 3 – Classificação das regiões em grupos de risco segundo valor da razão de frequência/expostos para cada cobertura.

Cobertura	Baixo Risco	Médio Risco	Alto Risco
Roubo/Furto + Incêndio	$[0; 0,005)$	$[0,005; 0,010)$	$[0,010, +\infty)$
Colisão parcial	$[0; 0,05)$	$[0,05; 0,10)$	$[0,10, +\infty)$
Colisão perda total	$[0; 0,005)$	$[0,005; 0,0075)$	$[0,0075, +\infty)$
Outras coberturas	$[0; 0,10)$	$[0,10; 0,15)$	$[0,15, +\infty)$

Assim, os critérios de corte para risco nas regiões foram alterados de forma a ter uma distribuição mais uniforme na quantidade de regiões classificadas em cada grupo¹⁷.

Além disso, as coberturas de colisão parcial e perda total foram desagrupadas, enquanto as coberturas de roubo/furto e incêndio foram fundidas, o que se detalha nas seções seguintes.

2.3.4 Desagrupamento das Coberturas de Colisão Parcial e Colisão Perda Total

Em virtude das inconsistências e problemas narrados sobre os dados de município e CEP, que prejudicam o uso nas análises, e considerando que somente os dados do arquivo segregado por sexo e faixa etária serão utilizados para fins de modelagem dos prêmios de risco, então é possível tratar as coberturas de colisão parcial e colisão perda total separadamente¹⁸. Ademais, as frequências e indenizações de colisão parcial e de colisão com perda total diferem sobremaneira entre si, sendo salutar modelá-las de forma distinta.

2.3.5 Agrupamento das Coberturas Roubo/Furto e Incêndio

Ao analisar os dados, observa-se que a frequência de sinistros de incêndio é bastante baixa, sendo que para diversos grupos de risco, fixado um modelo de veículo, sequer há dados de incêndio considerando todo o período (2012-2018). Além disso, os valores das indenizações por incêndio, para diversos modelos de veículos, assemelham-se às indenizações de roubo/furto.

Com isso, decidiu-se agrupar as variáveis resposta de frequência de sinistros e de indenização de sinistros de incêndio com os dados da cobertura de roubo/furto, modelando-os conjuntamente.

¹⁷ Ainda assim, é importante destacar que como o critério de classificação é único para todos os modelos de veículos, ocorre casos onde há somente duas regiões de risco com dados de sinistros para um modelo de veículo.

¹⁸ Isto só é possível no arquivo segregado por sexo e faixa etária, onde essas coberturas estão separadas. Nos arquivos por município e por CEP essas duas coberturas estão consolidadas numa única variável.

Este agrupamento também é realizado pela própria Susep no sistema de consultas *online* do Autoseg, onde essas coberturas só podem ser consultadas de forma agrupada.

2.4 Modelagem Estatística

Os fenômenos da natureza, sociais e econômicos costumam ser bastante complexos, dinâmicos e multivariados, sendo quase sempre regidos por leis inacessíveis ao ser humano. Explicar, de forma satisfatória, o comportamento dos diversos fenômenos que afligem o homem e a sociedade, ou ainda fazer previsões de eventos futuros, é uma tarefa difícil. Talvez a melhor forma de dar cor à compreensão dos fenômenos é por intermédio dos modelos matemáticos.

Modelos matemáticos, no fundo, nada mais são do que caricaturas da realidade. Em essência, um modelo matemático nunca conseguirá descrever, com perfeição, a verdade absoluta subjacente a um fenômeno observado. Ele pode, sim, aproximar-se com incrível exatidão em suas explicações e previsões, mas mesmo depois de ser testado por reiteradas vezes, e sempre se sair bem, ainda assim ele não poderá arvorar-se de ser um perfeito descritor da natureza¹⁹.

Nesse processo de explicar os fenômenos, invariavelmente se faz *associações* entre variáveis. Ao que parece, a necessidade de atribuir uma *causa* para tudo que ocorre na natureza é uma característica, ou uma angústia, puramente humana. Se chove hoje, as pessoas tentam buscar uma explicação, um motivo, para essa intempérie, indo da simples fúria de um deus desconhecido até a mais complexa combinação de variáveis, como umidade, temperatura, pressão atmosférica, ventos e época do ano. Alguns simplesmente atribuem o fenômeno ao *acaso*, em uma espécie de confissão – ou de rendição – perante a complexidade da natureza.

Enquanto os modelos determinísticos lutam ferozmente para estabelecer uma equação capaz de descrever um fenômeno, os modelos estatísticos notabilizam-se por darem voz ao acaso: neles, são combinadas possíveis *variáveis explicativas* de um fenômeno com um *fator estocástico*.

A literatura estabelece os seguintes princípios e etapas para a modelagem estatística de fenômenos (DOBSON, 2002), (MONTGOMERY; PECK; VINING, 2012), (TIBSHIRANI et al., 2013): análise exploratória dos dados; seleção ou elaboração do modelo; estimação de parâmetros; análise de resíduos e teste do modelo; previsão e interpretação.

A partir de constatações advindas da fase de análise exploratória, além de conhecimentos sobre o problema, parte-se para a seleção ou elaboração de um modelo. Obvi-

¹⁹ Talvez o exemplo mais emblemático seja a Mecânica Clássica, enunciada por Isaac Newton, testada em experimentos e observações por mais de 200 anos e que, anos mais tarde, descobriu-se ser inapropriada para explicar uma classe de fenômenos.

amente, é raro elaborar um modelo a partir de uma folha em branco; costumeiramente, nesta etapa, consulta-se a literatura ou a prática em determinado setor para selecionar modelos pré concebidos, documentados e de domínio da comunidade.

Pode-se dizer que as etapas de seleção do modelo, estimação de parâmetros, análise de resíduos e teste do modelo podem ser genericamente chamadas de *ajuste de modelo*, um processo que muitas das vezes é iterativo: não raro, depois de selecionar um modelo, estimar parâmetros e analisar resíduos, um modelo pode falhar enormemente em testes, motivando a revisão, ou repetição, de uma ou mais etapas desse processo. Ademais, um modelo sempre está limitado por suas hipóteses e aos dados sobre os quais foi ajustado; periodicamente, deve-se sempre colocá-lo à prova para checar sua satisfatoriedade e revisá-lo, se necessário.

Além dos princípios acima, também se deve mencionar o *princípio da parcimônia* (DOBSON, 2002). Segundo este princípio, um modelo mais simples, quando satisfatório, sempre é preferível em relação a modelos mais complexos (com mais variáveis), mesmo que estes últimos venham a ter um desempenho aparentemente melhor em suas previsões. Isso porque quanto mais simples um modelo é, tanto maior será a capacidade de interpretação e entendimento do fenômeno. Modelos com muitas variáveis explicativas ou muitos parâmetros, além de terem um maior custo para estimação de parâmetros e coleta de tantas variáveis, são difíceis de serem interpretados.

Por fim, principalmente em se tratando de modelagem estatística, *não existe modelo correto*. Por mais que a literatura teça recomendações ou haja consenso na adoção de determinadas modelagens para certos tipos de problema, o analista é livre para estabelecer qualquer modelo.

2.4.1 Estatística Bayesiana

A Estatística Bayesiana apresenta princípios e uma abordagem conceitualmente diferente da Estatística Clássica ou Frequentista, tema que será abordado brevemente nesta seção.

Na Estatística Clássica, todo e qualquer *parâmetro* presente em um modelo, seja um parâmetro indexador de uma função de probabilidade, ou por exemplo os coeficientes de um modelo de regressão, são tratados como se fossem *quantidades fixas*. Parâmetros são quantidades desconhecidas e não observáveis, mas sobre os quais precisamos formar conhecimento, de alguma maneira, para lançar mão das finalidades últimas dos modelos formulados: realizar *previsões* ou *interpretar fenômenos*.

Como os estatísticos clássicos consideram um parâmetro como uma quantidade fixa, um número desconhecido, eles necessitam *estimar* qual é o seu valor a fim de plugá-lo no modelo enunciado, uma vez que *descobrir* o seu verdadeiro valor não é possível. E, para isso, lançam mão de *estimadores*, que nada mais são do que funções de dados observáveis

que produzem *estimativas*, pontuais ou intervalares, para o parâmetro desconhecido. Para o estatístico clássico, a fonte de informação e conhecimento é única e exclusivamente proveniente dos dados observados – e com eles, o estatístico clássico faz todas as suas inferências.

Do outro lado, o estatístico bayesiano raciocina de forma distinta: “*Ora, se um parâmetro é uma quantidade desconhecida, então por que não dar tratamento probabilístico a ele?*” Esta é a primeira grande diferença da Estatística Bayesiana: dar *tratamento probabilístico* (ou de variável aleatória) a um parâmetro (CONGDON, 2005). Por mais que um parâmetro seja uma quantidade não observável, o estatístico bayesiano trata-o como se fosse uma variável aleatória, propondo um *modelo de probabilidade* para ele. A partir de crenças, experiências passadas, estudos anteriores ou informações de especialistas, os estatísticos bayesianos enunciam um modelo de probabilidade *a priori* para o parâmetro, ou seja, antes mesmo de observar qualquer dado. Quanto mais forte for a crença ou a experiência *a priori* do estatístico sobre o parâmetro, tanto mais ele traduzirá num modelo de alta precisão; se seu conhecimento *a priori* sobre o parâmetro for pouco ou nenhum, ele estabelecerá um modelo *vago*, com alta incerteza.

Depois de observar os dados, o estatístico bayesiano *atualiza* seus conhecimentos sobre o parâmetro, formando um novo conhecimento sobre ele *posteriormente* à observação dos dados. A informação proveniente dos dados, para o estatístico bayesiano, não serve para estimar diretamente um valor fixo de um parâmetro desconhecido, mas sim vai além: serve para que ele atualize as suas crenças probabilísticas sobre o parâmetro. Este processo de atualização é realizado por intermédio do *Teorema de Bayes* e é por isso que esse ramo da Estatística foi batizado com o nome de *Estatística Bayesiana*.

Em termos matemáticos, a distribuição *a posteriori* $\xi(\theta|x)$ para um parâmetro θ é proporcional ao produto da função de verossimilhança $l(\theta; \tilde{x})$ e a distribuição *a priori* $\xi(\theta)$ para θ , considerando que \tilde{x} é um vetor de dados observados (DEGROOT; SCHERVISH, 2012) e (MIGON; GAMERMAN, 1999).

$$\xi(\theta|x) \propto l(\theta; \tilde{x}) \cdot \xi(\theta). \quad (2.1)$$

Ainda que o paradigma bayesiano seja natural de ser compreendido, operacionalmente ele carrega uma dificuldade: a depender da quantidade de parâmetros e dos modelos de probabilidade em questão²⁰, encontrar a distribuição *a posteriori* para os parâmetros pode ser uma tarefa complexa, por envolver a resolução de integrais.

Contudo, graças ao imenso avanço da capacidade computacional nos últimos 50 anos, aliado à popularização dos computadores pessoais, é possível aceder à distribuição *a posteriori* sem ser necessário resolver analiticamente complexas integrais. A base da

²⁰ Incluindo a função de verossimilhança que foi enunciada.

inferência bayesiana consiste em obter *amostras de dados* da distribuição *a posteriori*, o que geralmente é realizado por intermédio de métodos iterativos de Monte Carlo via Cadeias de Markov (MCMC) (GAMERMAN; LOPES, 2006), (ROBERT; CASELLA, 2004) e (CONGDON, 2005). O desafio reside, no caso, em obter convergência nesses métodos iterativos, lançando mão de algoritmos eficientes para produzirem as amostras da distribuição *a posteriori*.

Com o tratamento probabilístico dado aos parâmetros e dispondo de amostras suficientemente grandes da distribuição *a posteriori*, ganha-se uma riqueza muito maior de informações sobre o parâmetro além de simples estimativas que a Estatística Clássica produz. Fazendo inferência bayesiana, pode-se responder perguntas como *qual a probabilidade de o parâmetro ser maior que tal valor* ou ainda produzir *intervalos de credibilidade* para o parâmetro, associados diretamente a uma probabilidade. Ademais, a lógica construtiva de um modelo bayesiano favorece sua revisão: à medida que novos dados chegam, a distribuição *a posteriori* atualiza-se, indo para uma nova distribuição. É possível mostrar que quanto mais dados chegam no modelo, mais a distribuição *a priori* dilui-se na verossimilhança, sendo que para grandes quantidades de dados observados, a distribuição *a posteriori* praticamente coincide com a verossimilhança.

Diante dessas vantagens e da riqueza de informações que fornece, este trabalho adotará o paradigma bayesiano para a construção dos modelos enunciados para precificação dos seguros de automóveis²¹.

2.5 Teoria da Credibilidade

O princípio do seguro alicerça-se na coletividade: indivíduos transferem seus riscos a uma seguradora mediante o pagamento de um prêmio, formando, assim, um grupo de segurados²². A coletividade é fundamental para a sustentabilidade de uma carteira de seguros pois alguns indivíduos, por força do acaso, serão acometidos por perdas enquanto outros não; é o montante de recursos aportados pelo grupo que possibilitará o pagamento das indenizações aos desafortunados.

Uma seguradora, quando decide assumir riscos, sabe que eles podem materializar-se no futuro em perdas com *certa probabilidade* e, para isso, necessita calcular o quanto deve cobrar para assumi-los. Eis uma das grandes dificuldades para as seguradoras, senão a principal: estimar essas probabilidades e precificar adequadamente qual quantia deverá

²¹ A Estatística Clássica possui como grande vantagem a rápida solução de problemas, quando estimadores com boas propriedades já foram estabelecidos e documentados. Pode-se dizer que seu “custo de implementação” é bastante menor em comparação à Estatística Bayesiana, mas paga-se o preço de não dispor de análises mais ricas. Como diz uma célebre frase, de autor desconhecido e reproduzida em diversos idiomas: “*Não existe almoço grátis na Matemática*”.

²² Risco nada mais é do que uma perda aleatória futura, a qual necessita ser traduzida em termos financeiros, de alguma forma. Até mesmo uma vida, considerada *sem preço*, precisa ser traduzida em termos monetários para fins de um seguro.

cobrar para assumir cada risco, em função dos indivíduos que compõem a carteira. Caso precise de forma inadequada, a seguradora corre o risco de tornar-se insolvente, se as perdas forem mais *frequentes* ou *severas* do que ela havia estimado.

Como a experiência individual de sinistros para cada apólice, na prática, costuma ser limitada e a quantidade de sinistros ocorridos geralmente é pequena, torna-se impraticável produzir estimativas precisas do prêmio considerando-se somente a experiência de cada indivíduo (BÜHLMANN; GISLER, 2005). Desta maneira, a saída é agrupar os indivíduos por similaridade de risco, construindo grupos homogêneos nos quais se acredita que todos comungam de um mesmo comportamento, facilitando a estimação de prêmios com maior precisão para o grupo todo, uma vez que a quantidade de indivíduos é maior. Ocorre que, se de um lado dispor de mais observações produz estimativas estatisticamente mais confiáveis, de outro lado à medida que indivíduos são agrupados para formarem grupos maiores, perde-se cada vez mais a homogeneidade entre eles. Desta maneira, as experiências individual e coletiva para cada risco devem ser combinadas para uma tarifação mais justa e adequada (BÜHLMANN; GISLER, 2005).

A Teoria da Credibilidade estuda formas de combinar a experiência individual com a experiência coletiva, com o intuito de estimar com mais qualidade os prêmios para um determinado risco. A ponderação atribuída à experiência individual é denominada *fator de credibilidade* (ω), dependendo do número de observações do indivíduo e da relação entre a variabilidade individual e a coletiva (BÜHLMANN; GISLER, 2005). A equação 2.2 apresenta como o prêmio individual (P^{ind}) e o prêmio coletivo (P^{col}) são combinados e ponderados pelo fator de credibilidade para resultar no prêmio de credibilidade (P^{cred}).

$$P^{\text{cred}} = \omega \cdot P^{\text{ind}} + (1 - \omega) \cdot P^{\text{col}} \quad (2.2)$$

O parâmetro θ designa um determinado perfil de risco, sendo que o prêmio é uma função desse perfil, ou seja, $\mu(\theta)$. Os valores dos sinistros em um determinado período de tempo²³ t são designados por X_t . Assim, o prêmio individual é definido pela equação 2.3:

$$P^{\text{ind}} = \mathbb{E}[X_{t+1}|\theta] =: \mu(\theta) \quad (2.3)$$

O interesse primordial está em produzir estimativas para $\mu(\theta)$ e uma das maneiras é por intermédio de estimadores. Como o prêmio individual a ser cobrado no instante de tempo $t + 1$ é dado pelo valor esperado de $X_{t+1}|\theta$ e, por ser uma esperança condicional, sabe-se que esta é uma variável aleatória, uma maneira quase que natural de estimar o prêmio individual seria tomando-se seu valor esperado, ou seja, $\mathbb{E}[\mathbb{E}[X_{t+1}|\theta]]$. No fundo,

²³ Usualmente, o período de tempo é anual.

quando se estima o prêmio individual dessa forma, o que se faz é tomar a média de todos os grupos de riscos de uma carteira; isto é, justamente, o prêmio coletivo (equação 2.4):

$$P^{\text{col}} = \mathbb{E}[\mathbb{E}[X_{t+1}|\theta]] = \mathbb{E}[\mu(\theta)] = \mu_0 \quad (2.4)$$

Ocorre que se o prêmio individual de cada grupo de risco fosse estimado pela equação 2.4, a experiência individual acabaria sendo ignorada e seu valor refletiria somente a média de todos os grupos de risco (que corresponde ao coletivo). Com isso, é necessário construir outras formas de estimar os prêmios individuais $\mu(\theta)$.

A partir da Teoria da Decisão, com respeito a uma função de perda quadrática (maiores detalhes ver em Bühlmann e Gisler), a estimativa de Bayes é considerada a melhor estimativa para o prêmio individual, sendo dada por:

$$P^{\text{Bayes}} = \mathbb{E}[\mu(\theta)|x_1, \dots, x_t] \quad (2.5)$$

A equação 2.5 é vista como a esperança *a posteriori* do prêmio individual. A obtenção do prêmio de credibilidade envolve o acesso à distribuição *a posteriori* de $\mu(\theta)|x_1, \dots, x_t$, o que nem sempre é fácil de ser realizado. Há maneiras de aceder ao valor esperado da distribuição *a posteriori* por aproximação linear de $\mathbb{E}[\mu(\theta)|x_1, \dots, x_t]$ via modelos lineares de credibilidade, especificando somente alguns parâmetros sob determinadas hipóteses e construindo estimadores, sem estabelecer modelos de probabilidade *a priori* para todos os parâmetros relacionados ao perfil de risco θ , conforme proposto por Bühlmann (BÜHLMANN, 1965).

Uma forma de ilustrar como as equações 2.2, 2.3, 2.4 e 2.5 relacionam-se entre si é por meio de um exemplo. Assumindo que, condicionado a um perfil de risco θ , os sinistros X_t são independentes e identicamente distribuídos segundo uma distribuição normal, tem-se que: $X_t|\theta \stackrel{iid}{\sim} \mathcal{N}(\mu(\theta), \sigma^2)$ e que a média $\mu(\theta)$ (para cada perfil de risco) tem uma distribuição normal $\mu(\theta) \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \tau^2)$.

Sob essas hipóteses e aplicando as definições das equações 2.3 e 2.4 nos modelos normais enunciados do exemplo, tem-se que (BÜHLMANN; GISLER, 2005):

$$P^{\text{ind}} = \mu(\theta) \quad (2.6)$$

$$P^{\text{col}} = \mu_0 \quad (2.7)$$

Segundo a equação 2.1, conjuga-se os dois núcleos das duas distribuições normais (da verossimilhança e da *priori* para $\mu(\theta)$) a fim de aceder ao núcleo da distribuição *a posteriori* para $\mu(\theta)|x_1, \dots, x_n$; em seguida, toma-se o valor esperado *a posteriori* $\mathbb{E}[\mu(\theta)|x_1, \dots, x_n]$ – o que corresponde ao estimador de Bayes sob uma função de perda

quadrática (P^{Bayes}). Aplicando esse procedimento ao exemplo em discussão, é possível mostrar que:

$$P^{\text{Bayes}} = \frac{\tau^2 \mu_0 + \sigma^2 \bar{X}_n}{\tau^2 + n\sigma^2} \quad (2.8)$$

onde $\bar{X}_n = \sum_{t=1}^n \frac{X_t}{n}$ representa a média de sinistros em n anos num determinado perfil de risco θ . Tomando-se ω como sendo:

$$\omega = \frac{n}{n + \frac{\sigma^2}{\tau^2}} \quad (2.9)$$

é possível reescrever a equação 2.8 como a ponderação entre prêmio individual e prêmio coletivo tal qual estabelecida na 2.2. A média de sinistros em um determinado perfil de risco é um estimador para $\mu(\theta)$. Ou seja, $\hat{\mu}(\theta) = \bar{X}_n$ e tem-se:

$$\begin{aligned} P^{\text{Bayes}} &= \omega \cdot \bar{X}_n + (1 - \omega) \cdot \mu_0 \\ &= \omega \cdot P^{\text{ind}} + (1 - \omega) \cdot P^{\text{col}} \end{aligned} \quad (2.10)$$

Contudo, no exemplo acima, os parâmetros σ^2 e τ^2 não tiveram distribuições *a priori* especificadas no modelo. Estimadores poderiam ser desenvolvidos para essas quantidades, mas uma maneira mais rica para solucionar o problema é realizar a completa especificação de distribuições para os parâmetros, conforme introduzido na subseção a seguir.

2.5.1 Modelo de Bühlmann-Straub

O modelo de Bühlmann-Straub surge como uma extensão ao modelo de Bühlmann. No modelo de Bühlmann, em uma carteira com $i = 1, \dots, m$ grupos de risco, dispõe-se de observações ao longo do tempo $t = 1, \dots, n_i$ para cada um dos grupos de risco θ_i , sendo o sinistro denotado por X_{it} (BÜHLMANN; GISLER, 2005).

As hipóteses para o modelo de Bühlmann são:

1. Carteira de seguros balanceada: cada grupo de risco θ_i possui a mesma quantidade de unidades expostas ao risco.
2. $X_{it}|\theta_i$ são independentes e identicamente distribuídos, com distribuição caracterizada por $\mathbb{E}[X_{it}|\theta_i] = \mu(\theta_i)$ e $\text{Var}[X_{it}|\theta_i] = \sigma^2(\theta_i)$, para $i = 1, \dots, m$ riscos;
3. $\mu(\theta_i) \sim \xi(\theta_i)$ com $\mathbb{E}[\mu(\theta_i)] = \mu_0$ e $\text{Var}[\mu(\theta_i)] = \tau^2$.

Segundo o modelo de Bühlmann, os indivíduos pertencentes a um mesmo grupo de risco θ_i são supostos similares entre si e o parâmetro $\sigma^2(\theta_i)$ representa a variância dos

sinistros dentro de um determinado grupo. Já o parâmetro τ^2 é interpretado como sendo a variância entre os grupos de risco, bem como o parâmetro μ_0 representa um nível médio de sinistro comum a todos os grupos, ou seja, equivalente ao prêmio coletivo.

A hipótese de carteira balanceada no modelo de Bühlmann é bastante forte e nem sempre é observada, o que motiva uma alteração deste modelo. Surge, então, o modelo de Bühlmann-Straub (BÜHLMANN; STRAUB, 1970), o qual insere uma medida de volume de unidades expostas ao risco (v_{it}) para cada grupo θ_i (BÜHLMANN; GISLER, 2005).

No modelo de Bühlmann-Straub, utiliza-se o valor do sinistro agregado S_{it} ponderado por uma unidade de volume v_{it} , ambos associados ao grupo de risco i no período de tempo t , fazendo com que $X_{it} = \frac{S_{it}}{v_{it}}$. Assim, no fundo, tem-se uma taxa de prêmio advinda de uma razão de perda, ou uma razão de sinistro. As demais hipóteses do modelo de Bühlmann são mantidas idênticas no modelo de Bühlmann-Straub.

A completa especificação do modelo de Bühlmann-Straub é realizada ao se estabelecer modelos de probabilidade para todos os parâmetros e quantidades aleatórias envolvidos. Uma forma de compreender como é feita a completa especificação de um modelo de Bühlmann-Straub é por intermédio de um exemplo, o que é apresentado na próxima seção com o caso real do estudo do presente trabalho.

2.6 Modelagem para o Protótipo

As subseções a seguir apresentam as especificações dos modelos de severidade dos sinistros, frequência de sinistros e a combinação entre frequência-severidade.

É importante destacar que para cada uma das coberturas de seguro²⁴ é ajustado um modelo separadamente dos demais, presumindo-se que cada cobertura comporta-se de maneira independente.

Também, foi assumido que cada um dos 17 modelos de veículo do presente estudo corresponde a uma carteira. Ou seja, quando se fala em prêmio coletivo, fala-se do prêmio que engloba todos os veículos de um determinado modelo, sendo os grupos de risco construídos a partir da combinação de variáveis de um mesmo modelo de veículo, como: subgrupo do modelo, idade do veículo, sexo do condutor, sua faixa etária e região de circulação. Logo, a quantidade m de grupos de risco varia de acordo com cada um dos 17 modelos de veículos e também para cada cobertura, em função dos agrupamentos propostos.

Os modelos de Bühlmann-Straub propostos neste trabalho devem ser interpretados tendo-se em mente que o modelo de veículo e a cobertura estão fixados. Com isso, considerando que o protótipo contempla 17 modelos de veículos, cada um deles com 4

²⁴ Roubo/furto e incêndio, colisão parcial, colisão perda total e outras coberturas.

coberturas e que há duas variáveis de interesse (frequência e severidade), foram ajustados 136 modelos de Bühlmann-Straub para os dados de 2012-2018, sendo que as amostras das distribuições *a posteriori* para os parâmetros de cada modelo foram obtidas por simulação via MCMC utilizando-se a biblioteca RStan.

2.6.1 Modelo para Severidade

Inicialmente, define-se a razão de indenizações X_{it} como sendo a relação entre o total de indenizações S_{it} ponderada pela frequência de sinistros v_{it} em determinado grupo de risco e ano. Ou seja, $X_{it} = \frac{S_{it}}{v_{it}}$ representa a severidade de um sinistro, dado que ele ocorreu.

A partir da análise exploratória realizada, e de forma generalizada para os diversos modelos de veículos e coberturas estudadas, constatou-se que a distribuição $X_{it}|\theta_i$ possui aspecto assimétrico positivo, quando fixado um modelo de veículo e uma cobertura, sugerindo a adoção de modelos que deem conta dessa assimetria.

Tomando-se a variável $Y_{it}|\theta_i = \log(X_{it}|\theta_i)$, realizaram-se testes de hipóteses de Shapiro-Wilk de normalidade sobre a variável $Y_{it}|\theta_i$, de forma a testar se há evidências suficientes para rejeitar o modelo normal para $Y_{it}|\theta_i$. Os resultados dos testes²⁵ conduziram amplamente para a não rejeição da normalidade para $Y_{it}|\theta_i$, o que leva à decisão de que não há evidências fortes o suficiente para concluir que a adoção de um modelo *lognormal* para $X_{it}|\theta_i$ não seja plausível. Assim, considerou-se que: $X_{it}|\theta_i \sim \mathcal{LN}(\mu(\theta_i), v_{it}^{-1}\sigma^2(\theta_i))$ e, de forma equivalente, $Y_{it}|\theta_i \sim \mathcal{N}(\mu(\theta_i), v_{it}^{-1}\sigma^2(\theta_i))$.

Na análise exploratória, viu-se que a variância dentro de cada grupo de risco $\sigma^2(\theta_i)$ aparenta ter diferenças quando se compara os diversos grupos. Contudo, não se consegue extrair com facilidade sobre como cada grupo de risco está associado com essas diferenças de variância (dentre os diversos veículos e as diversas coberturas), o que dificulta sobremaneira uma modelagem que considere variâncias distintas em cada grupo de risco²⁶. Com isso, assumiu-se que $\sigma^2(\theta_i) = \sigma^2$, ou seja, que a variância é a mesma para todos os grupos de risco. Dentro do modelo, a variância aparece multiplicada pelo inverso do volume de unidades expostas em cada risco i e ano t observado, fazendo com que grupos de menor volume tenham maior incerteza.

Um modelo de Bühlmann-Straub para $Y_{it}|\theta_i = \log(X_{it}|\theta_i)$ poderia ser assim estabelecido:

²⁵ Ao nível de significância de 5%.

²⁶ Para modelar cada grupo de risco com variância distinta, adiciona-se um novo vetor paramétrico ao modelo, de dimensão igual à quantidade de grupos de risco em cada cobertura, que tem a finalidade de governar sobre como é o aspecto da variância de cada grupo. Ocorre que há determinados modelos de veículos, para determinadas coberturas, que há dezenas de grupos de risco distintos, o que dificultaria sobremaneira a especificação de tantas constantes em tal vetor.

$$Y_{it}|\theta_i \sim \mathcal{N}(\mu(\theta_i), v_{it}^{-1}\sigma^2) \quad (2.11)$$

$$\mu(\theta_i)|\mu_0, \tau^2 \sim \mathcal{N}(\mu_0, \tau^2) \quad (2.12)$$

$$\mu_0 \sim \mathcal{N}(a_0, b_0^2) \quad (2.13)$$

$$\sigma \sim \mathcal{U}[0, k_0] \quad (2.14)$$

$$\tau \sim \mathcal{U}[0, k_0] \quad (2.15)$$

Ocorre que nesse modelo enunciado presume-se que a média da severidade $\mu(\theta_i)$, para cada grupo de risco, é oriunda de uma mesma distribuição de probabilidade sem qualquer dependência temporal, o que não parece ser razoável para o problema em estudo. A análise exploratória revelou que, com o passar dos anos e para diversos modelos de veículos, grupos de risco e coberturas, a severidade tem uma aparente tendência de crescimento com o passar do tempo, muito provavelmente justificada pela inflação sobre serviços de reparos, peças automotivas e alterações macroeconômicas que afetam o valor dos veículos²⁷.

A maneira de ponderar a experiência observada ao longo do período de tempo $T = 1, \dots, t$ a fim de determinar o prêmio em $t + 1$ (como, por exemplo, dando mais peso às observações mais recentes e menos peso às observações passadas), motiva a elaboração de novos modelos que desempenhem esse papel. Uma forma usual de modelar o prêmio de risco no instante de tempo $t + 1$ é utilizando-se de um sistema de *Bonus-Malus*, onde o prêmio P_{t+1} leva em conta o prêmio de risco praticado no tempo anterior (P_t), agregando-se a ele a experiência X_t observada no tempo t (geralmente, associada ao número de sinistros) ajustada por um peso ou medida de credibilidade α_t (BÜHLMANN; GISLER, 2005). Assim, tem-se uma fórmula recursiva de cálculo do prêmio na forma da equação 2.16.

$$P_{t+1} = P_t + \alpha_t(X_t - P_t) \quad (2.16)$$

Ocorre que o prêmio individual e o peso α_t da equação 2.16 podem mudar com o passar do tempo sob uma complexa dinâmica. O prêmio individual é afetado por uma série de covariáveis²⁸ de intrincada associação entre elas, o que motiva o lançamento de modelos em que se dê voz a *fatores estocásticos*, os quais evoluem com o passar do tempo por intermédio de equações de evolução (BÜHLMANN; GISLER, 2005).

²⁷ Muitas apólices de seguro auto são indexadas pelo valor do veículo segundo a Tabela Fipe, que nada mais é do que uma avaliação média do carro a valores de mercado, sujeita, portanto, ao cenário econômico do país. É bastante usual que veículos depreciem com o passar do tempo mas, nos recentes anos, fenômeno contrário tem sido observado: os veículos têm mantido seu valor ou até mesmo se valorizado (UOL, 2021).

²⁸ Muitas delas sequer são possíveis de serem medidas pelas seguradoras, como uma mudança de emprego, o nascimento de um filho, divórcio etc., as quais podem afetar o risco do segurado de uma forma bem particular, além de aspectos coletivos que se alteram com o tempo, como mudanças no trânsito ou na violência urbana de determinados locais.

Uma série de diferentes abordagens para levar em conta a evolução temporal dos prêmios é apresentada por Bühlmann e Gisler (2005) nos denominados *modelos de credibilidade evolutivos*, incluindo o desenvolvimento de estimadores para cada caso. Neste trabalho, considerando-se a adoção de uma abordagem bayesiana, propõe-se uma simples modificação ao modelo de Bühlmann-Straub anteriormente enunciado, fazendo com que a média $\mu(\theta_i)$ passe a ter influência temporal e também visualize o nível do ano anterior, tornando-se $\mu_t(\theta_i)$. A introdução de um fator de crescimento $\beta_t(\theta_i)$, dependente do tempo t e do grupo de risco i , possibilita que a média da severidade atualize-se com um comportamento de aumento ou queda, com o passar dos anos, sendo possível também fazer estimativas sobre $\beta_t(\theta_i)$. Ademais, o fator de crescimento também tem sua variância ponderada pelo volume v_{it} , fazendo com que haja maior incerteza sobre $\beta_t(\theta_i)$ em grupos de menor volume, enquanto em grupos volumosos haja maior precisão sobre ele.

O modelo de Bühlmann-Straub modificado é descrito pelas seguintes relações:

$$Y_{it}|\mu_t(\theta_i), \sigma^2 \sim \mathcal{N}(\mu_t(\theta_i), v_{it}^{-1}\sigma^2) \quad (2.17)$$

$$\mu_t(\theta_i) = \mu_{t-1}(\theta_i) + \beta_t(\theta_i) \quad (2.18)$$

$$\beta_t(\theta_i)|\beta_0, \tau^2 \sim \mathcal{N}(\beta_0, v_{it}^{-1}\tau^2) \quad (2.19)$$

$$\mu_0(\theta_i) \sim \mathcal{N}(a_0, b_0^2) \quad (2.20)$$

$$\beta_0 \sim \mathcal{N}(0, c_0^2) \quad (2.21)$$

$$\sigma \sim \mathcal{U}[0, k_0] \quad (2.22)$$

$$\tau \sim \mathcal{U}[0, k_0] \quad (2.23)$$

O valor esperado de $\mu_t(\theta_i)|y_{i1}, \dots, y_{it}$ é dado por:

$$\begin{aligned} \mathbb{E}[\mu_t(\theta_i)|y_{i1}, \dots, y_{it}] &= \mathbb{E}[(\mu_{t-1}(\theta_i) + \beta_t(\theta_i))|y_{i1}, \dots, y_{it}] \\ &= \mathbb{E}[\mu_{t-1}(\theta_i)|y_{i1}, \dots, y_{it-1}] + \mathbb{E}[\beta_t(\theta_i)|y_{i1}, \dots, y_{it}] \end{aligned} \quad (2.24)$$

À exceção dos modelos de veículos *Fiat Argo*, *Jeep Compass* e *Renault Kwid*, os demais veículos possuem observações disponíveis para o período de 2012-2018 nos diversos grupos de risco e, com isso, preferiu-se adotar distribuições vagas *a priori* para os parâmetros²⁹ $\mu_0(\theta_i)$, β_0 , σ e τ . Para as *priori* vagas, considerou-se que $a_0 = 0$ e $b_0 = c_0 = k_0 = 10^6$.

Para os modelos de veículos *Fiat Argo*, *Jeep Compass* e *Renault Kwid*, há somente uma única observação referente ao ano de 2018, pois estes modelos de veículos eram recém

²⁹ As simulações foram todas realizadas sobre a variável $Y_{it}|\theta_i = \log(X_{it}|\theta_i)$.

lançados à época. Com isso, especificamente para eles, foi necessário introduzir crenças *a priori* sobre os parâmetros $\mu_0(\theta_i)$, β_0 , σ e τ as quais, naturalmente, variam em função de cada modelo de veículo e também da cobertura (pois a severidade é associada ao valor do carro).

Inicialmente, reduziu-se a incerteza sobre β_0 , σ e τ ao estabelecer $c_0 = 1$ e $k_0 = 30$, com conhecimento advindo da modelagem de outros modelos de veículos. Quanto aos valores de a_0 e b_0 , que controlam a distribuição *a priori* para $\mu_0(\theta_i)$, aplicou-se um procedimento que leva em conta os valores de comercialização dos veículos 0 km no mês de referência de dezembro/2018 (FIPE, 2018), uma vez que as coberturas contra roubo/furto, incêndio e colisão com perda total geralmente levam ao pagamento de indenizações da ordem do valor do veículo; já para colisão parcial e outras coberturas, especulou-se que as indenizações girariam em torno de 20% a 7%, respectivamente, dos valores dos veículos 0 km. Os parâmetros a_0 e b_0 das distribuições *a priori* para $\mu_0(\theta_i)$ estão sumarizados na Tabela 4.

Tabela 4 – Parâmetros a_0 e b_0 da distribuição *a priori* para $\mu_0(\theta_i)$.

Cobertura	Modelo	a_0	b_0
Roubo/furto + Incêndio	Fiat Argo	10,76	0,061
	Jeep Compass	11,86	0,061
	Renault Kwid	10,38	0,061
Colisão parcial	Fiat Argo	9,15	0,135
	Jeep Compass	10,25	0,135
	Renault Kwid	8,77	0,135
Colisão perda total	Fiat Argo	10,76	0,061
	Jeep Compass	11,86	0,061
	Renault Kwid	10,38	0,061
Outras coberturas	Fiat Argo	8,10	0,135
	Jeep Compass	9,20	0,135
	Renault Kwid	7,72	0,135

A partir da geração de amostras *a posteriori* para Y_{it} , o retorno às indenizações ponderadas por unidade de frequência é realizado por $X_{it}|\theta_i = \exp(Y_{it}|\theta_i)$.

O fator de credibilidade ω_i para cada grupo de risco i pode ser calculado pelo rearranjo da equação 2.2:

$$\omega_i = \frac{P_i^{\text{cred}} - P_i^{\text{col}}}{P_i^{\text{ind}} - P_i^{\text{col}}} \quad (2.25)$$

2.6.2 Modelo para Frequência

Considerando a distribuição Poisson para o número de sinistros N_{it} , condicionando-se a um determinado grupo de risco θ_i , admite-se que os $N_{it}|\theta_i$ são independentes, ou seja, fixado um determinado grupo de risco, a quantidade de sinistros entre os anos observados é independente³⁰. Como a distribuição Poisson caracteriza-se por ter média e variância iguais ao valor de seu parâmetro indexador, tem-se que:

$$\mathbb{E}[N_{it}|\theta_i] = \text{Var}[N_{it}|\theta_i] = \lambda_{it}^* = v_{it}\lambda(\theta_i) = v_{it}\lambda_0\theta_i \quad (2.26)$$

A reparametrização acima indica o seguinte: a taxa anual da frequência de sinistros, em um grupo de risco i e no ano t , é controlada pelo volume de unidades expostas ao risco v_{it} , por um nível λ_0 comum a todos os grupos de risco e por um parâmetro θ_i que depende do grupo de risco considerado.

O modelo para a frequência é especificado da seguinte forma:

$$N_{it} \sim \text{Poi}(\lambda_{it}) \quad (2.27)$$

$$\lambda_{it} = v_{it}\lambda(\theta_i) = v_{it}\lambda_0\theta_i \quad (2.28)$$

$$\theta_i \sim \Gamma(\theta_0, \theta_0) \quad (2.29)$$

$$\lambda_0 \sim \Gamma(a_0, b_0) \quad (2.30)$$

Da mesma forma que no modelo de severidade, os veículos *Fiat Argo*, *Jeep Compass* e *Renault Kwid* tiveram crenças *a priori* inseridas sobre os parâmetros θ_i e λ_0 , as quais diferenciam-se para cada cobertura³¹, uma vez que esses veículos só possuem observação para o ano de 2018. Os valores de a_0 e b_0 para suas *priori* estão apresentados na Tabela 5; para estes veículos, também se reduziu a variância *a priori* para θ_i , tomando-se $\theta_0 = 0,1$.

Os demais modelos de veículos foram modelados com distribuições vagas *a priori* para os parâmetros θ_i e λ_0 , tomando-se $\theta_0 = a_0 = b_0 = 0,001$.

A partir da frequência de sinistros N_{it} e da quantidade de veículos expostos v_{it} em cada ano e grupo de risco, tem-se qual a frequência de sinistros por unidade de volume $F_{it} = \frac{N_{it}}{v_{it}}$, sendo que as amostras *a posteriori* para $\lambda(\theta_i)$ indicam o prêmio de credibilidade para a frequência ponderada pelas unidades expostas a risco.

³⁰ Em outras palavras, neste modelo, o fato de haver muitos sinistros no ano t não condiciona o que irá ocorrer no ano $t + 1$, sendo t e $t + 1$ dois anos dentro do período observacional dos dados.

³¹ Utilizado conhecimento *a priori*, obtido dos demais modelos de veículos, sobre qual a localização e dispersão da frequência de sinistros ponderada pela unidade de volume, em cada cobertura. P.ex., as frequências de colisão parcial ou do acionamento da assistência 24 horas costumam ser bem maiores que as frequências de colisão com perda total ou de roubo/furto.

Tabela 5 – Parâmetros a_0 e b_0 da distribuição *a priori* para λ_0 dos veículos *Fiat Argo*, *Jeep Compass* e *Renault Kwid*.

Cobertura	a_0	b_0
Roubo/furto + Incêndio	1	100
Colisão parcial	1	10
Colisão perda total	2	500
Outras coberturas	3	20

O fator de credibilidade ω_i para cada grupo de risco i é calculado pela equação 2.31:

$$\omega_i = \frac{v_i}{v_i + \frac{\theta_0}{\lambda_0}} \quad (2.31)$$

Da mesma forma que ocorre no modelo para severidade, grupos de risco de menor volume possuirão menor fator de credibilidade, ou seja, o prêmio coletivo entra com maior peso no cálculo dos prêmios de credibilidade.

2.6.3 Modelo para Frequência-Severidade

Como o escopo do protótipo é apresentar estimativas de prêmios a um interessado em seguro, o interesse último reside na estimação do prêmio frequência-severidade, pois é ele que constituirá o prêmio de risco considerado para fins de precificação a um segurado, trazendo uma medida da razão entre o montante esperado de sinistros para cada unidade exposta ao risco. Em termos simples:

$$\frac{\text{Sinistros}}{\text{Expostos}} \stackrel{[=]}{=} \frac{\text{Frequência}}{\text{Expostos}} \cdot \frac{\text{Sinistros}}{\text{Frequência}} \quad (2.32)$$

Formalizando a equação 2.32, a partir dos prêmios de credibilidade obtidos das amostras *a posteriori* para a frequência ponderada pelo volume $\lambda(\theta_i)$ e para a severidade $\mu_t(\theta_i)$, o prêmio de frequência-severidade para um determinado grupo de risco i pode ser determinado por:

$$P_t^{\text{cred}}(\theta_i) = \mathbb{E}[\lambda(\theta_i)] \cdot \mathbb{E}[\mu_t(\theta_i)], \quad \text{para } i = 1, \dots, m \quad (2.33)$$

A equação 2.33 reproduz o modelo do risco coletivo anual enunciado na Teoria do Risco, onde se utiliza o conceito de risco agregado em uma carteira (FERREIRA, 2010). No caso do presente trabalho, as previsões estão sendo realizadas para o ano $t = 2019$. Para o cálculo do prêmio de frequência-severidade na equação 2.33, adotam-se as amostras das distribuições *a posteriori* simuladas para o ano $t = 2018$. Para a frequência, tem-se que

$\mathbb{E}[\lambda(\theta_i)|y_{i2012}, \dots, y_{i2018}]$ (distribuição *a posteriori* no ano de 2018) desempenha a previsão para o ano de 2019. Já para o caso da severidade, e por intermédio da equação 2.24, faz-se com que a distribuição *a posteriori* para $\mu_t(\theta_i)|y_1, \dots, y_t$ e $\beta_t(\theta_i)|y_1, \dots, y_t$ em $t = 2018$ desempenhem papel de distribuição *a priori* em $t = 2019$, somando-as para se ter uma previsão para a severidade no ano de 2019.

Destaca-se que a média de uma distribuição *a posteriori* corresponde ao estimador de Bayes com respeito a uma função de perda quadrática (DEGROOT; SCHERVISH, 2012). Nada impede, todavia, que se tomasse a mediana *a posteriori*³² ou ainda quantis *a posteriori* (como 95% ou 99%) para construção das estimativas do prêmio de frequência-severidade.

³² Estimador de Bayes com respeito a uma função de perda linear.

3 Resultados e Discussão

Este capítulo está dividido em três seções: a primeira seção apresenta as constatações principais da fase de análise exploratória; a segunda contém a descrição dos principais resultados alcançados com os modelos estatísticos ajustados; e, por fim, a última seção faz uma discussão global sobre os aspectos considerados mais relevantes na modelagem proposta e resultados alcançados.

3.1 Módulo Análise Exploratória

Os resultados da análise exploratória estão disponibilizados no *app* Shiny em <https://rbivanov.shinyapps.io/autocheck-TCC-AED/>. As constatações foram construídas mediante análises dos gráficos, aliadas a conhecimentos existentes sobre o mercado de seguros automóveis, e estão descritas nas subseções a seguir. A Figura 1 apresenta uma das visualizações do módulo de análise exploratória, com mapas geográficos construídos para análise das variáveis em função do município de localização.

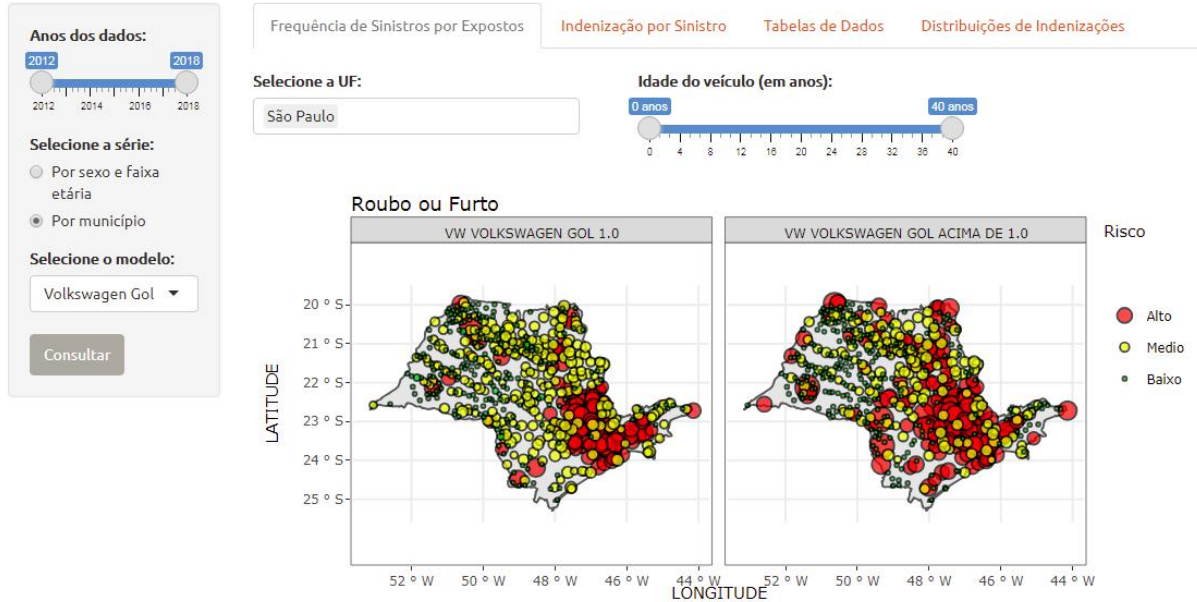


Figura 1 – Visualização do módulo de análise exploratória do *app* Autocheck.

É importante salientar que em função da quantidade de modelos de veículos em estudo e também da quantidade de variáveis analisadas, as constatações podem diferir em alguns modelos de veículos, subgrupos de modelos, anos analisados etc., sendo que a análise descrita a seguir apresenta o cenário majoritário das constatações. Também, como a modelagem estatística proposta será única para todos os modelos de veículos em estudo,

buscou-se uniformizar o entendimento para todos eles o máximo possível, a exceção do *Fiat Argo*, *Jeep Compass* e *Renault Kwid*, que por serem veículos novos e com poucos dados disponíveis, tiveram um tratamento diferenciado.

Para formação de um entendimento preliminar dos dados de frequência e indenizações, os dados foram analisados considerando-se todo o período (2012-2018) e todas as UF do Brasil, ponderando-se sempre os dados pela quantidade de unidades expostas ao risco.

3.1.1 Cobertura Roubo ou Furto

A frequência de roubos e furtos por unidades expostas possui uma associação aparentemente forte com o município de circulação, constatado em todos os modelos de veículos, o que faz sentido: espera-se que em municípios maiores (ou com maiores índices de criminalidade) haja maiores índices de roubo e furto de veículos. Para o subgrupo do modelo, observou-se que em alguns casos há uma associação maior deste na frequência de roubo/furto, indicando prováveis preferências dos criminosos por determinadas especificações de carros. Também, ainda que de forma mais fraca, foi visto que a frequência de roubos/furtos tende a decair sutilmente com o aumento da faixa etária do condutor.

Para o sexo do condutor e a idade do veículo, foi observado, em alguns modelos, uma aparente associação dessas variáveis com a frequência de roubos e furtos, enquanto que para outros modelos não há uma clareza nessa associação. De toda forma, essas variáveis aparentaram uma associação bem mais fraca com a frequência de roubos do que aquilo que se observa em relação ao município ou subgrupo do modelo.

Para o valor de indenização de um roubo/furto, dado que um sinistro ocorreu, nota-se uma influência aparentemente maior do subgrupo do modelo, sendo que a idade do veículo também impacta principalmente no tocante à variabilidade dos valores observados.

Desta maneira, as seguintes variáveis foram identificadas como candidatas a variáveis explicativas para a cobertura de roubo/furto:

- Frequência: município, subgrupo do modelo e faixa etária do condutor;
- Severidade: subgrupo do modelo e idade do veículo.

3.1.2 Cobertura Colisão Parcial ou Colisão Perda Total

A frequência de colisões ponderada pelas unidades expostas possui uma associação bastante clara com o sexo e faixa etária do condutor, observado nos diversos modelos, tanto para perda parcial como para perda total. De forma geral, há uma frequência maior de colisões entre os mais jovens, além de que, dada uma mesma faixa etária, os condutores do sexo masculino aparentam ter uma maior frequência de colisões, em média, do que as

condutoras do sexo feminino. Também, o subgrupo do modelo aparentou ter uma associação com a frequência de colisões, provavelmente por conta das diferenças de potência dos veículos.

O município de circulação também exerce uma aparente influência na frequência de colisões, provavelmente pela organização do trânsito de cada cidade e cultura dos condutores. Curiosamente, em uma análise preliminar, não aparenta haver uma associação clara entre a frequência de colisões somente com o fato de o município ser populoso: há casos onde municípios menores, populacionalmente e com menos expostos, tiveram uma frequência de colisões proporcionalmente maior que em capitais populosas.

Para o valor de indenização, dado que uma colisão ocorreu, também se notou uma influência maior do subgrupo do modelo e da idade do veículo, em condição análoga ao observado para o caso de roubo/furto.

Assim, foram identificadas as seguintes variáveis para as duas coberturas de colisão:

- Frequência: faixa etária, sexo do condutor, subgrupo do modelo e município;
- Severidade: subgrupo do modelo e idade do veículo.

3.1.3 Cobertura Incêndio

A cobertura de incêndio mostrou-se como a mais complexa de colher *insights* na análise exploratória, principalmente pelo fato de haver muitos grupamentos onde a frequência de sinistros (e conseqüentemente de indenizações) foi igual a zero.

Para os modelos que possuem mais de um subgrupo do modelo, há uma sutil e aparente diferença em média e variabilidade para a frequência e indenizações de incêndios quando comparados os diferentes subgrupos de modelos, fazendo com que essa fosse a única variável considerada para fins explicativos da frequência e severidade dos incêndios nos diversos modelos, conjuntamente com a idade do veículo.

Conforme justificado na metodologia, a cobertura de incêndio foi incorporada à cobertura de roubo/furto, tanto para modelagem de frequência quanto severidade.

3.1.4 Outras Coberturas

As outras coberturas, que englobam assistência 24 h, vidros, blindagem, equipamentos acessórios etc., apresentaram um comportamento bastante peculiar: da análise exploratória, nota-se que há um aparente aumento na frequência desses sinistros principalmente dentre as condutoras do sexo feminino e, principalmente, em idades mais avançadas (notadamente, acima dos 55 anos de idade). Em contrapartida, as indenizações (dado que um sinistro ocorreu) não acompanham o aumento dessa frequência maior de sinistros para as mulheres acima de 55 anos de idade: os valores médios das indenizações aparentam ser

menores, dando indícios de que as mulheres com mais idade acionam mais o seguro para essas coberturas, mas o valor das indenizações não necessariamente são altos. O município também aparenta estar associado à frequência de sinistros das outras coberturas.

Com isso, as seguintes variáveis foram elecandas para a modelagem dessa cobertura:

- Frequência: faixa etária, sexo do condutor e município;
- Severidade: faixa etária, sexo do condutor, subgrupo do modelo e idade do veículo.

3.1.5 Resumo das Constatações

A Tabela 6 sintetiza as variáveis identificadas na análise exploratória para modelagem da frequência e severidade das diversas coberturas. Para fins de construção dos modelos estatísticos, o município foi substituído pelo grupo de risco da região de circulação e a idade do veículo por seu grupo de risco definido.

Tabela 6 – Variáveis identificadas na análise exploratória para modelagem da frequência e severidade.

Cobertura	Frequência	Severidade
Roubo/furto + Incêndio	Município Subgrupo do modelo Faixa etária	Subgrupo do modelo Idade do veículo
Colisão parcial	Faixa etária Sexo Subgrupo do modelo Município	Subgrupo do modelo Idade do veículo
Colisão perda total	Faixa etária Sexo Subgrupo do modelo Município	Subgrupo do modelo Idade do veículo
Outras coberturas	Faixa etária Sexo Município	Faixa etária Sexo Subgrupo do modelo Idade do veículo

Com o agrupamento realizado para a variável de *idade do veículo* e a adoção da variável *região de circulação* (em substituição ao *município*) também com um agrupamento, as variáveis selecionadas para a modelagem possuem a seguinte quantidade de níveis:

- Subgrupo do modelo: 1 a 4 níveis (a depender do modelo);

- Sexo do condutor: 2 níveis;
- Faixa etária do condutor: 5 níveis;
- Risco da região: 3 níveis;
- Grupo de idade do veículo: 3 níveis.

Ou seja, a depender do modelo do veículo, de qual cobertura e qual variável resposta (se frequência ou se severidade) está realizando a modelagem, haverá uma quantidade distinta de grupos de risco para cada caso.

Por exemplo, conforme a Tabela 6, a modelagem da frequência de roubo/furto do *Fiat Uno* envolve 60 grupos de risco distintos, uma vez que há 3 níveis para a região de circulação, 4 níveis para o subgrupo do modelo e 5 níveis para a faixa etária do condutor.

Cada uma dessas combinações, em todos os casos, foi considerada como sendo um *grupo de risco* para fins do ajuste dos modelos de credibilidade.

3.1.6 Codificação dos Grupos de Risco

Em função das coberturas e das variáveis resposta (frequência ou severidade), há um conjunto de variáveis explicativas distinto, onde cada uma delas possui certos níveis, os quais compõem-se para formar um grupo de risco.

Assim, foi adotado um procedimento para codificar e identificar um grupo de risco no formato GXYZRRI, de acordo com a seguinte regra e na seguinte ordem:

1. G : letra que significa “grupo”;
2. $X \in \{1, 2, 3, 4\}$: identifica o subgrupo do modelo do veículo (ver Apêndice B);
3. $Y \in \{F, M\}$: sexo do condutor (feminino ou masculino);
4. $Z \in \{1, 2, 3, 4, 5\}$: faixa etária do condutor (respectivamente, de 18 a 25 anos; 26 a 35 anos; 36 a 45 anos; 46 a 55 anos; e acima de 55 anos de idade);
5. $RR \in \{AR, MR, BR\}$: risco da região de circulação (respectivamente, alto, médio e baixo risco);
6. $I \in \{A, B, C\}$: grupo da idade do veículo (respectivamente, de 0 a 3 anos; 4 a 10 anos; e 11 a 40 anos).

Por exemplo, o código de risco G2F3MR para o *Ford Fiesta* indica que se trata do grupo *Ford Fiesta acima de 1.0*, do sexo feminino, com faixa etária de 36 a 45 anos de idade e numa região de médio risco. O código G1A para o *Volkswagen Gol* designa o *Volkswagen Gol 1.0* que possui de 0 a 3 anos de idade¹.

¹ O número que vem imediatamente depois da letra G sempre indicará o código do subgrupo do modelo.

3.2 Módulo Precificação

A partir da modelagem descrita na seção 2.6 e implementada no R, os resultados produzidos foram alimentados no *app* [Autocheck - Precificação](#), onde o usuário pode consultar, por intermédio de abas e menus, as estimativas de prêmios produzidas para o ano de 2019 e compará-las com as reais observações daquele ano, para um determinado perfil de segurado construído a partir das seguintes variáveis: modelo, subgrupo do modelo, ano do veículo, sexo, faixa etária e região de circulação do condutor.

Além disso, é possível verificar uma série de informações sobre o modelo estatístico ajustado, como as estatísticas-resumo das amostras *a posteriori* e informações sobre a convergência das amostras geradas por simulação via MCMC.

As subseções a seguir apresentam uma discussão sobre os principais resultados observados com a modelagem estatística adotada.

AutoCheck - Protótipo para Precificação de Seguros de Automóvel

Elaborado por Ricardo Boeira Ivanov (EIBA 427)
Rio de Janeiro, 10 de agosto de 2022.

Módulo: Precificação (TCC)

Seja bem-vindo ao Autocheck, uma ferramenta para você estimar o preço de seguro do seu automóvel a partir de dados públicos da Superintendência de Seguros Privados (Susep). Antes de realizar sua consulta, entenda melhor como a ferramenta foi desenvolvida e funciona, acessando a aba de apresentação.

Figura 2 – Visualização do *app* Autocheck para precificação de seguros.

3.2.1 Principais Resultados

O cerne do Autocheck encontra-se na aba “*Prêmio do Seguro*”, onde é exibido o prêmio de credibilidade de frequência-severidade, conforme a seleção de perfil do usuário, incluindo os resultados reais do ano de 2019 para fins comparativos (ver Figura 2).

Já a aba “*Parâmetros do Modelo*” apresenta estatísticas-resumo² das amostras *a posteriori* para $\lambda(\theta_i)$ (modelo de frequência) e $\mu_t(\theta_i)$ (modelo de severidade, para $t = 2018$) de todos os grupos de risco para um modelo de veículo selecionado, incluindo também

² Mínimo, primeiro quartil, mediana, média, terceiro quartil e máximo.

os intervalos de credibilidade bilaterais com 95% de probabilidade para cada um desses parâmetros, além da observação do ano de 2019 para fins comparativos. Nesta aba, a navegação ocorre selecionando-se a cobertura de seguro desejada.

As abas “*Frequência*” e “*Severidade*” apresentam gráficos das amostras *a posteriori* geradas por simulação via MCMC para os diversos parâmetros dos modelos ajustados, incluindo também gráficos de autocorrelação empírica para checar se é razoável assumir que as amostras obtidas possam ser assumidas como sendo independentes. Essas abas ilustram as amostras para o perfil selecionado pelo usuário. Estas abas, em conjunto com a aba “*Parâmetros do Modelo*”, têm finalidade acadêmica e fazem parte somente do *app* Autocheck acessório ao presente trabalho, não sendo intenção mantê-las num *app* aos usuários gerais.

Por fim, é importante destacar que caso o usuário faça uma seleção específica cujo grupo de risco, porventura, não possuía nenhuma observação no período 2012-2018 (e, portanto, não foi possível prever um prêmio de credibilidade para o ano de 2019), o Autocheck apresenta o *prêmio coletivo* para o grupo de risco selecionado. Na mesma esteira, quando um determinado grupo de risco selecionado pelo usuário não possui observações para o ano de 2019, o Autocheck exhibe valores zerados para o ano.

3.2.2 Avaliação do Modelo

As Tabelas 7 e 8 apresentam os prêmios estimados de frequência-severidade e as indenizações observadas no ano de 2019, respectivamente, por modelo de veículo e por cobertura de seguro, possibilitando o cálculo da sinistralidade para cada carteira.

O detalhamento de cada cobertura nos modelos de veículos encontra-se nas Tabelas 11 e 12, constantes nos Apêndices C e D, respectivamente. Já os resultados obtidos para os grupos de risco dentro de um determinado modelo de veículo e cobertura estão disponíveis no [Autocheck - Precificação](#).

Também no Autocheck é possível conferir os intervalos de credibilidade para as amostras das distribuições *a posteriori* dos prêmios de frequência e de severidade no ano de 2018, as quais servem de previsão para o ano de 2019, permitindo visualizar se esses intervalos de credibilidade contêm, ou não, o valor das observações do ano de 2019. As Tabelas 9 e 10 sumarizam a taxa de acertos dos intervalos de credibilidade considerando todos os grupos de risco simulados³, indicando também se o modelo superestimou ou subestimou o valor dos prêmios nos casos em que o intervalo de credibilidade não conteve a observação do ano de 2019. Nas referidas tabelas, erros com sinal (>) indicam a quantidade de previsões erradas superestimadas; símbolos (<) indicam a quantidade de previsões erradas aquém do intervalo de credibilidade.

³ Recapitulando que, dado um modelo de veículo e cobertura, o grupo de risco é definido conforme

Tabela 7 – Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, sintetizados por modelo de veículo.

Modelo	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Chevrolet Celta	117.848.439,15	102.592.785,00	0,87
Chevrolet Corsa	156.117.223,47	144.861.335,00	0,93
Chevrolet Onix	428.153.106,45	366.190.184,00	0,86
Chevrolet Prisma	217.285.646,64	215.301.557,00	0,99
Fiat Argo	51.403.975,38	41.691.552,00	0,81
Fiat Palio	331.682.205,41	306.527.942,00	0,92
Fiat Siena	215.225.740,68	209.069.657,00	0,97
Fiat Strada	215.796.293,97	195.067.080,00	0,90
Fiat Uno	205.057.993,61	176.472.526,00	0,86
Ford Fiesta	304.130.477,43	279.893.274,00	0,92
Ford Ka	224.222.964,07	229.290.489,00	1,02
Hyundai HB20	607.327.983,08	456.137.182,00	0,75
Jeep Compass	123.007.845,80	115.467.586,00	0,94
Renault Kwid	46.736.145,26	43.712.432,00	0,94
Volkswagen Fox	392.177.839,29	342.175.427,00	0,87
Volkswagen Gol	339.438.277,70	290.772.362,00	0,86
Volkswagen Polo	102.560.879,34	86.447.032,00	0,84
Total	4.078.173.036,73	3.601.670.402,00	0,88

Tabela 8 – Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, sintetizados por cobertura.

Cobertura	Prêmio (R\$)	Indenizações (R\$)	Sinistralidade
Roubo/Furto + Incêndio	1.411.517.160,32	1.025.599.570,00	0,73
Colisão Parcial	1.332.571.142,15	1.124.891.792,00	0,84
Colisão Perda Total	855.492.103,14	878.793.438,00	1,03
Outras coberturas	478.592.631,12	572.385.602,00	1,20
Total	4.078.173.036,73	3.601.670.402,00	0,88

3.3 Discussão Global dos Resultados

De antemão, pode-se afirmar que a modelagem proposta foi bastante satisfatória para tarifar os prêmios de risco para o ano de 2019 dos veículos em estudo, a partir dos dados públicos do Autoseg.

Para todas as carteiras construídas a sinistralidade em 2019 concentrou-se majoritariamente em torno de 1,00, denotando que as previsões situaram-se próximas aos reais valores das indenizações que viriam a ocorrer em 2019.

descrito na subseção 3.1.6.

Tabela 9 – Desempenho das previsões para a frequência $\lambda(\theta_i)$ considerando intervalos de credibilidade bilaterais com 95% de probabilidade.

Cobertura	Qtd. Riscos	Taxa de Acertos	Erros
Roubo/Furto + Incêndio	448	22%	319 (>) 29 (<)
Colisão Parcial	765	10%	675 (>) 11 (<)
Colisão Perda Total	770	43%	206 (>) 230 (<)
Outras coberturas	503	14%	98 (>) 335 (<)

Tabela 10 – Desempenho das previsões para a severidade $\mu_t(\theta_i)$ considerando intervalos de credibilidade bilaterais com 95% de probabilidade.

Cobertura	Qtd. Riscos	Taxa de Acertos	Erros
Roubo/Furto + Incêndio	77	99%	1 (>) - (<)
Colisão Parcial	77	87%	- (>) 10 (<)
Colisão Perda Total	76	96%	- (>) 3 (<)
Outras coberturas	734	89%	31 (>) 48 (<)

Por se tratar de prêmio de risco, é importante destacar que a sinistralidade, neste caso, deve situar-se em torno da unidade, sem estar muito distante dela, seja aquém ou além. Isto porque o prêmio de risco, por definição, corresponde ao *valor esperado* dos sinistros e, assim, prêmio de risco e indenizações observadas devem aproximar-se em um bom processo de estimação.

As subseções a seguir relatam alguns pontos principais sobre o processo de modelagem estatística adotado e as previsões realizadas.

3.3.1 Modelo para Frequência de Sinistros

A taxa de acertos para o modelo de frequência de sinistros, quando se leva em conta os intervalos de credibilidade, foi consideravelmente baixa, conforme se vê na Tabela 9. Acredita-se que o baixo desempenho do modelo de frequência (sob o critério dos intervalos de credibilidade) deve-se a duas causas principais.

Diferentemente do modelo proposto para severidade, o modelo de frequência de sinistros não possui qualquer dependência temporal para o parâmetro $\lambda(\theta_i)$. A única influência temporal que ataca o parâmetro, controlador da taxa de ocorrência de sinistros, é o volume de unidades expostas a risco no grupo de risco i e no ano t considerado (v_{it}). Excetuando pelo volume, os demais parâmetros do modelo são considerados identicamente distribuídos sob o aspecto temporal. Ou seja, admite-se que a taxa $\lambda(\theta_i)$ no horizonte de 2012 a 2018, fixado um grupo de risco i , é sempre a mesma. Neste sentido, caso um novo modelo para frequência de sinistros seja estabelecido levando-se em conta a influência temporal, acredita-se que haja uma maior exatidão e precisão nas previsões realizadas⁴.

Em segundo lugar, ressalta-se que o modelo de frequência foi construído sobre um modelo de *Poisson*, onde a média e variância são iguais. Ocorre que a frequência de sinistros costuma ser baixa, puxando a reboque a variância para valores baixos. Assim, os intervalos de credibilidade têm uma menor amplitude, o que aumenta as chances de não incluírem a observação do ano de 2019 em suas previsões. Uma alternativa para contornar essa situação seria a proposição de uma modelagem alternativa ao modelo *Poisson* usual adotado.

Coincidentemente, o modelo de frequência na cobertura de roubo/furto e incêndio acabou *superestimando* a ocorrência desses sinistros, enquanto o modelo nas outras coberturas (assistência 24 horas etc.) *subestimou* sua frequência. No cômputo de todas as coberturas em uma apólice de um mesmo veículo, um efeito acabou compensando o outro e graças aos bons resultados da modelagem para a severidade, o prêmio frequência-severidade para um veículo apresentou bons resultados. Registra-se, assim, que a modelagem de frequência de sinistros dessas duas coberturas pode ser aprimorada, levando a melhores resultados especificamente para elas.

3.3.2 Modelagem de Veículos Novos

Os modelos *Fiat Argo*, *Jeep Compass* e *Renault Kwid* possuem uma única observação no horizonte observacional (referente ao ano de 2018) por terem sido recém lançados naquela época. A fim de não inflar demasiadamente a incerteza das previsões para o ano de 2019, foi necessário lançar mão de distribuições *a priori* carregadas de crenças sobre os parâmetros.

Os parâmetros para as distribuições *a priori* foram construídos a partir de conhecimento adquirido mediante a observação do comportamento dos parâmetros *a posteriori* nos outros modelos de veículos simulados, agregando-se a informação do preço dos carros segundo a tabela Fipe em 2018 para fins de posicionamento dos parâmetros no caso da severidade.

⁴ Neste trabalho, por razões de eficiência dos algoritmos e do volume de modelos a simular, não foi realizada esta vinculação temporal no caso da frequência.

Pelo fato de haver uma única observação disponível para os veículos novos, as amostras *a posteriori* dos diversos parâmetros apresentam uma dispersão naturalmente superior quando comparadas às amostras dos demais modelos simulados, aumentando, por consequência, a amplitude dos intervalos de credibilidade. Também, com uma única observação, a incerteza *a posteriori* fica calcada nas crenças *a priori*.

Na modelagem dos veículos novos, preservou-se o pressuposto basal da inferência bayesiana: não “espiar” os dados para ajustar uma distribuição *a priori* mais condizente com eles e nem voltar ao modelo, trocando a distribuição *a priori* por outra, em decorrência de uma insatisfação com os resultados obtidos *a posteriori*.

Mesmo assim, analisando os resultados agregados das carteiras do *Fiat Argo*, *Jeep Compass* e *Renault Kwid*, seus resultados podem ser considerados satisfatórios. Registra-se que um estudo mais aprofundado sobre como elicitare distribuições *a priori* para modelagem de frequência e severidade em veículos novos pode (e deve) ser realizado.

3.3.3 Fenômeno de Nascimento, Migração e Morte de Grupos de Risco

De acordo com a análise exploratória, notou-se que a idade do veículo tem uma associação com o valor das indenizações, sendo que carros mais novos levam a indenizações mais altas enquanto carros antigos possuem menor valor de indenização⁵. Portanto, para formar os grupos de risco na modelagem da severidade, a idade do veículo foi considerada e agrupada em três categorias: Grupo A (veículos 0 km com até 3 anos de uso); Grupo B (veículos com 4 a 10 anos de uso); e Grupo C (veículos com mais de 10 anos de uso).

Ocorre que alguns dos modelos de veículos estudados iniciaram sua fabricação e comercialização dentro do período de 2012 a 2018, ou poucos anos antes, fazendo com que os Grupos B ou C ainda não tivessem nenhuma unidade exposta a risco nos anos iniciais. Com o passar do tempo, os Grupos B e C *nascem* a partir do momento que um modelo veículo passa a ter mais do que 4 e 10 anos de uso, respectivamente. Há, portanto, uma *migração* do volume de unidades expostas a risco com o passar do tempo⁶.

Também, alguns modelos de veículos pararam de ser fabricados durante o período de 2012 a 2018, fazendo com que o Grupo A, paulatinamente com o passar dos anos, reduzisse o volume de unidades expostas até zero, ocasionando a sua *morte*. Quando um veículo para de ser fabricado, os Grupos A e B morrem depois de 4 e 10 anos a partir da interrupção de fabricação de um determinado modelo, restando tão-somente o Grupo C⁷.

⁵ Isso está associado com a avaliação do veículo na tabela Fipe, indexador de praticamente todas as apólices de veículos no Brasil.

⁶ O nascimento de grupos de risco foi observado no *Chevrolet Onix* e *Chevrolet Prisma*, e em determinados subgrupos de modelo do *Fiat Siena*, *Fiat Uno* e *Volkswagen Fox*.

⁷ No estudo realizado, o *Chevrolet Celta* e *Chevrolet Corsa* pararam de ser fabricados em 2016 e 2015, respectivamente, havendo a morte do Grupo A em 2019 e 2018.

Para a finalidade do Autocheck, a morte de um grupo de risco não constitui um problema, uma vez que quando um grupo deixa de existir ele sequer necessita ser modelado, pois não haverá essa seleção ao usuário. O problema reside no *nascimento* de um grupo de risco: quando um grupo de risco nasce no meio do período observacional (2012-2018), isso faz com que haja menos observações para aquele grupo de risco específico, comparando com os demais grupos de risco. Se o nascimento de um grupo ocorre só em 2018, o vetor observacional não possui dados de 2012 a 2017, levando a um problema para obtenção da convergência nas simulações, visto que os grupos de risco comungam de parâmetros comum a todos eles (como β_0 , σ e τ), o que fatalmente infla a variância. Quando o nascimento do grupo ocorreu nos primeiros anos ou em algum ano intermediário do período observacional, constatou-se que o fator de crescimento $\beta_t(\theta_i)$ conseguiu administrar a falta de dados anteriores de forma razoável, não afetando a convergência das cadeias.

No caso de veículos que tiveram o nascimento de um novo grupo de risco ao fim do período observacional, houve problema de convergência nas cadeias dos parâmetros associados à variância, necessitando assim uma medida corretiva e intervenção pontual na modelagem. Nesses casos, seria possível adotar-se uma distribuição *a priori* carregada de informação somente para esse grupo de risco específico, de forma similar ao desenvolvido para os veículos novos. Todavia, como dentro de um modelo de veículo costuma haver vários grupos de risco distintos, preferiu-se adotar o valor do *prêmio coletivo* para as previsões do prêmio nesses casos.

A problemática de nascimento e morte de grupos de risco, por conta da idade do veículo, deixaria de existir caso a avaliação dos veículos, segundo a tabela Fipe, fosse introduzida na modelagem em substituição à idade do veículo. Ainda assim, o fenômeno de nascimento e morte de grupos de risco constitui-se somente em uma dificuldade adicional na construção e simulação dos modelos, não repercutindo em falhas de conclusão ou de obtenção de resultados satisfatórios.

3.3.4 Covariáveis ausentes e não utilizadas

As seguradoras dispõem de informações bastante relevantes para estimar a frequência e severidade de sinistros, as quais não estão disponíveis na base pública do Autoseg, tais como: se o segurado utiliza o veículo para se deslocar ao trabalho ou faculdade, se possui estacionamento na residência ou trabalho, se utiliza o veículo diretamente para trabalhar com ele (caso de motoristas), existência de itens opcionais (como kit GNV) etc. Acredita-se que a presença dessas covariáveis na base do Autoseg e sua introdução nos modelos estatísticos aumentaria o poder das previsões realizadas.

Quanto à localização das apólices, conforme amplamente dissertado na seção 2.2.3.3, acredita-se que a solução das inconsistências cadastrais de municípios e CEP no Autoseg possibilitaria a construção de uma modelagem espacial de maior robustez (CZADO; GS-

CHLÖSSL, 2007), aumentando enormemente o poder preditivo da frequência de sinistros de roubo/furto.

4 Conclusões

O objetivo pretendido por este trabalho foi alcançado: precificar prêmios de risco para seguros de automóvel a partir de dados públicos e disponibilizar um protótipo de aplicativo para consulta das informações.

De forma geral, os modelos de credibilidade bayesianos de Bühlmann-Straub ajustados com dados de 2012-2018 foram capazes de prever de forma bastante satisfatória o valor esperado das indenizações para o ano de 2019 nos 17 modelos de veículos estudados, demonstrando o poder desses modelos.

A abordagem bayesiana trouxe riqueza às análises realizadas, possibilitando a construção de intervalos de credibilidade e realização de previsões de forma mais inteligível, facilitando também o processo de modificação do modelo de Bühlmann-Straub (no caso da severidade) sem a necessidade de ter que deduzir complexas expressões matemáticas para estimadores e investigar suas propriedades, o que nem sempre é uma tarefa trivial. Também, a abordagem bayesiana possibilitou o contorno a questões de escassez de observações no caso da modelagem de veículos novos, permitindo a combinação de crenças externas aos dados para produzir melhores estimativas.

Conclui-se que a modelagem estatística proposta e o Autocheck como ferramenta para consulta de prêmios têm potencial para serem estendidos a todos os veículos segurados no Brasil, entregando à sociedade informações relevantes que auxiliem os cidadãos no processo de avaliação de preços das seguradoras e no conhecimento da ciência atuarial por trás da tarifação de seguros.

Sugestões para Trabalhos Futuros

Além da extensão do Autocheck para todos os modelos de veículos, com a posterior construção e disponibilização do aplicativo em larga escala, uma série de outros trabalhos podem ser desenvolvidos a partir do presente estudo. A lista de sugestões a seguir não é exaustiva e não está em ordem de prioridade ou sugestão.

- Realizar modelagem espacial para a frequência de sinistros de roubo/furto considerando o município e o CEP da apólice;
- Desenvolver metodologia para elicitar distribuições *a priori* para veículos novos;
- Introduzir na modelagem estatística a avaliação do valor dos veículos segundo tabela Fipe;
- Desenvolver um pacote no R que possibilite a leitura dos dados do Autoseg diretamente no terminal do R;
- Aprimorar a eficiência dos algoritmos para simulação via MCMC com alternativas ao uso do RStan;
- Introduzir a influência temporal na modelagem da frequência de sinistros;
- Desenvolver métodos formais para agrupar e classificar as variáveis que definem um grupo de risco;
- Estender o presente estudo para veículos cujo volume de unidades expostas a risco é baixo (como veículos de luxo, importados etc.) e também para motocicletas, caminhões e ônibus.

Referências

- AUTOO. *Carros mais vendidos do Brasil em 2018*. Autoo. Dez. 2019. Disponível em: <https://www.autoo.com.br/emplacamentos/veiculos-mais-vendidos/2018/>. Acesso em: 4 jan. 2022.
- AUTOPAPO. *Entenda o que é proteção veicular e suas diferenças para o seguro*. UOL. Fev. 2021. Disponível em: <https://autopapo.uol.com.br/noticia/o-que-e-protecao-veicular/>. Acesso em: 24 jun. 2022.
- AZEVÊDO, Beatriz Margarida Zanotto de. *Impactos causados por diferentes tipos de modelos de veículos na precificação de seguros auto na cidade de São Paulo utilizando modelos GAMLSS e o método de simulações de Monte Carlo*. 2020. Trabalho de Conclusão de Curso em Ciências Atuariais – Escola Paulista de Política, Economia e Negócios, Universidade de São Paulo.
- BÜHLMANN, H. Experience Rating and Credibility. *Astin Colloquium*, Lucerne, 1965. Disponível em: <https://www.cambridge.org/core/journals/astin-bulletin-journal-of-the-iaa/article/experience-rating-and-credibility/B9A879CE4A73397C653963B474A7F954>.
- BÜHLMANN, H.; GISLER, A. *A Course in Credibility Theory and its Applications*. Springer, 2005.
- BÜHLMANN, H.; STRAUB, E. Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Bern, 1970.
- CESAR, Brian Santos Furtado. *Avaliação da ocorrência de sinistros através da aplicação de algoritmos de machine learning*. 2020. Trabalho de Conclusão de Curso em Ciências Atuariais – Escola Paulista de Política, Economia e Negócios, Universidade de São Paulo.
- CONGDON, P. *Bayesian Models for Categorical Data*. John Wiley & Sons, Ltd, 2005.
- CZADO, C.; GSCHLÖSSL, S. Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, Taylor & Francis, v. 3, p. 202–225, 2007. Disponível em: <http://dx.doi.org/10.1080/03461230701414764>.
- DEGROOT, M.H.; SCHERVISH, M.J. *Probability and Statistics*. 4th edition: Pearson Education Inc., 2012.
- DOBSON, A.J. *An Introduction to Generalized Linear Models*. 2. ed.: Chapman & Hall/CRC texts in statistical science series, 2002.
- FARIAS, T.A.; JESUS, J.C. Tarificação de Seguros de Automóveis no Mercado do Nordeste. *RIC Revista de Informação Contábil*, v. 15, p. 1–16, 2019.

- FERREIRA, P.P. *Modelos de Precificação e Ruína para Seguros de Curto Prazo*. 2ª edição: Fundação Escola Nacional de Seguros (Funenseg), 2010.
- FIPE. *Preço médio de veículos: consulta de carros e utilitários pequenos*. Fipe. Dez. 2018. Disponível em: <https://veiculos.fipe.org.br/>. Acesso em: 20 jul. 2022.
- GAMERMAN, D.; LOPES, H.F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2nd edition: Chapman e Hall/CRC, 2006.
- MIGON, H.S.; GAMERMAN, D. *Statistical Inference: an Integrated Approach*. 1st edition: A Hodder Arnold Publication, 1999.
- MINFRA. *Frota de Veículos - mensal - 03-2020*. Dados extraídos do Registro Nacional de Veículos Automotores (Renavam). Ministério da Infraestrutura. Mar. 2020. Disponível em: <https://dados.infraestrutura.gov.br/dataset/frota-de-veiculos/resource/67d403c2-22cd-4f4b-b75e-650143d37141>. Acesso em: 25 fev. 2022.
- MINFRA. *Frota Registrada de Veículos por UF, tipo e modelo - 10-2021*. Ministério da Infraestrutura. Out. 2021. Disponível em: <https://www.gov.br/infraestrutura/pt-br/assuntos/transito/arquivos-senatran/estatisticas/renavam/2021/outubro/>. Acesso em: 25 fev. 2022.
- MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. *Introduction to Linear Regression Analysis*. 5th edition: Wiley, 2012.
- POLICYGENIUS. *Car insurance statistics in 2022*. Policygenius. Mar. 2022. Disponível em: <https://www.policygenius.com/auto-insurance/car-insurance-statistics/>. Acesso em: 23 jun. 2022.
- PRAZERES, Bárbara Ivânia Rangel. *Definição de grupos de risco de roubo ou furto, perda parcial ou total de veículos via k-means*. 2020. Trabalho de Conclusão de Curso em Ciências Atuariais – Escola Paulista de Política, Economia e Negócios, Universidade de São Paulo.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.
- ROBERT, C.P.; CASELLA, G. *Monte Carlo Statistical Methods*. 2nd edition: Springer, 2004. Disponível em: <https://link.springer.com/book/10.1007/978-1-4757-4145-2>.
- STATISTA. *Car insurance in the U.S. - statistics & facts*. Statista. Jan. 2022. Disponível em: <https://www.statista.com/topics/3087/car-insurance-in-the-united-states/>. Acesso em: 23 jun. 2022.
- SUSEP. *Autoseg - Sistema de Estatísticas de Automóveis da Susep*. Superintendência de Seguros Privados. Jun. 2020. Disponível em: <http://www2.susep.gov.br/menuestatistica/Autoseg/principal.aspx>. Acesso em: 15 jan. 2022.

- SUSEP. *SES - Sistema de Estatísticas da Susep*. Superintendência de Seguros Privados. Dez. 2021. Disponível em: http://www2.susep.gov.br/menuestatistica/SES/premio_sesinistros.aspx?id=54. Acesso em: 11 mar. 2022.
- TEIXEIRA, F.J.; SCALON, J.D. Análise espacial do prêmio médio do seguro de automóvel em Minas Gerais. *Revista da Estatística UFOP*, Universidade Federal de Ouro Preto, v. 3, 2014.
- TIBSHIRANI, R. et al. *An Introduction to Statistical Learning with Applications in R*. 1st edition: Springer, 2013.
- UOL. *Carros usados têm valorização de até 27%*. UOL. Set. 2021. Disponível em: <https://noticias.uol.com.br/videos/?id=carros-usados-tem-valorizacao-de-ate-27-04028D993868DCC96326>. Acesso em: 4 ago. 2022.
- UOL. *Como fazer seguro de carro mais velho? Veja as dicas*. UOL. Out. 2018. Disponível em: <https://www.uol.com.br/carros/noticias/redacao/2018/10/24/como-fazer-seguro-de-carro-mais-velho-veja-as-dicas.htm>. Acesso em: 24 jun. 2022.
- UOL. *Veja ranking dos 10 carros usados mais vendidos no Brasil em 2018*. UOL. Jan. 2019. Disponível em: <https://www.uol.com.br/carros/noticias/redacao/2019/01/14/veja-ranking-dos-10-carros-usados-mais-vendidos-no-brasil-em-2018.htm>. Acesso em: 4 jan. 2022.
- ZANIBONI, N.; MONTINI, A. Modelos de Poisson inflada de zeros e binomial negativa inflada de zeros na previsão de sinistros de automóveis. *E&G Economia e Gestão*, PUC Minas, v. 15, 2015.

Apêndices

APÊNDICE A – Exemplos de Gráficos da Análise Exploratória

Os gráficos deste Apêndice apresentam exemplos não exaustivos da análise exploratória, acompanhados de uma breve discussão sobre os achados em cada um deles e como influenciou a fase de modelagem.

Na Figura 3 é possível identificar, para o *Volkswagen Gol*, que a frequência média de colisões parciais ou com perda total são aparentemente distintas em função do sexo e da faixa etária do condutor, já que as caixas estão posicionadas em alturas distintas nos gráficos quando se altera os níveis dessas variáveis. Os condutores do sexo masculino aparentam ter uma frequência de colisões superior às condutoras do sexo feminino, quando fixada uma idade. Adicionalmente, para ambos os sexos, a frequência média de colisões aparentam ter uma redução à medida que os condutores ficam mais velhos. Também, as duas janelas dos gráficos (uma para cada subgrupo do modelo do veículo) mostram que há aparente diferença média quando se troca o subgrupo do modelo de veículo.

Essas constatações direcionam para a inclusão das variáveis sexo, faixa etária e subgrupo do modelo do veículo para a construção dos grupos de risco de frequência de colisões parciais e por perda total.

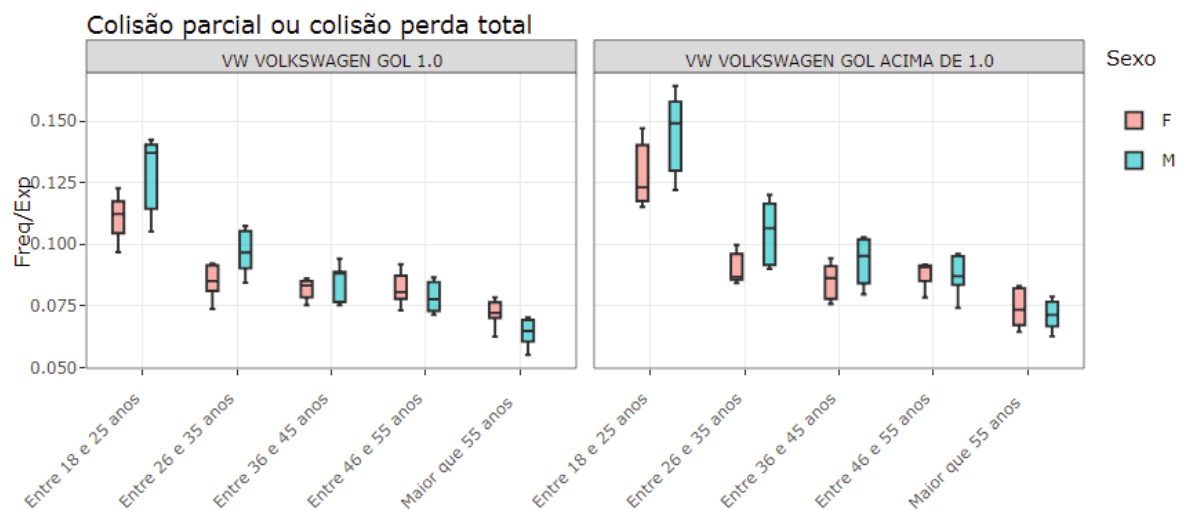


Figura 3 – *Boxplot* da frequência de colisão parcial ou perda total do *Volkswagen Gol* ponderada por unidades expostas em função do sexo, faixa etária do condutor e subgrupo do modelo (seleção Brasil e todos os anos de modelo).

A Figura 4 apresenta as indenizações por roubo e furto do *Chevrolet Corsa* ponde-

radas pela exposição, segregadas por sexo, faixa etária e subgrupo do modelo. Dado que um sinistro ocorreu, é difícil constatar de que forma o sexo e a faixa etária afetam nas indenizações médias por roubo e furto, mas é bem claro que o subgrupo de modelo têm uma aparente influência nessa média (uma vez que o valor do veículo está intimamente ligado ao subgrupo do modelo). Desta maneira, indica-se para a seleção da variável de subgrupo do modelo para a construção dos grupos de risco da cobertura de roubo ou furto.

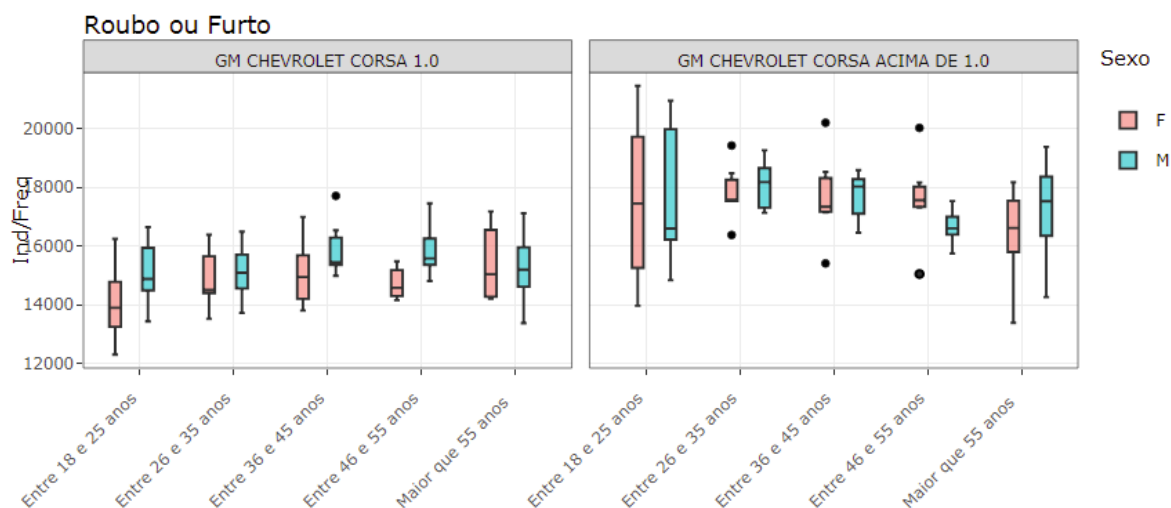


Figura 4 – *Boxplot* das indenizações de roubo ou furto do *Chevrolet Corsa* ponderada por unidades expostas em função do sexo, faixa etária do condutor e subgrupo do modelo (seleção Brasil e todos os anos de modelo).

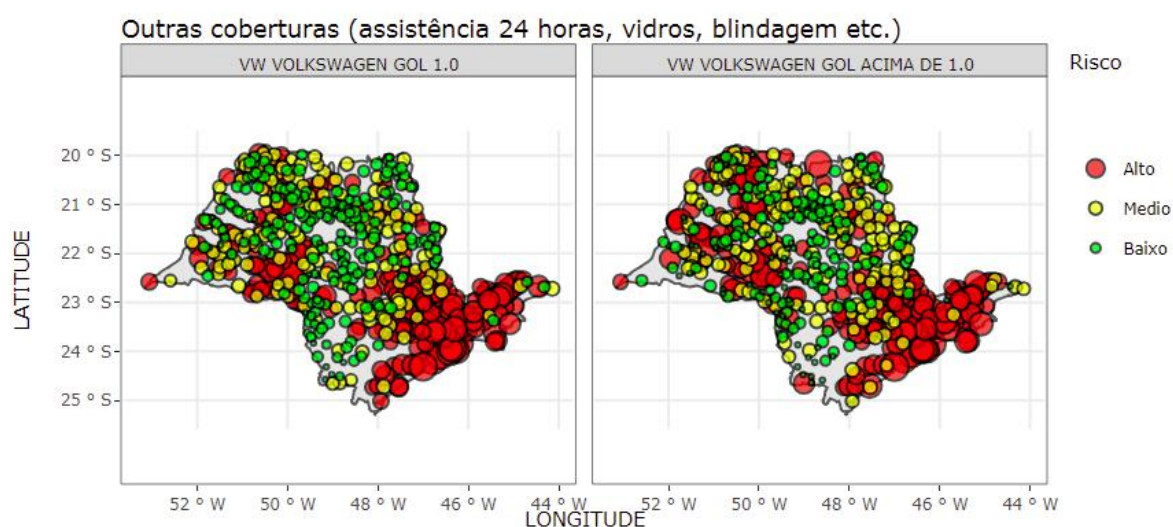


Figura 5 – Mapa do estado de São Paulo da frequência de outras coberturas do *Volkswagen Gol* ponderada por unidades expostas em função do subgrupo do modelo e município.

Os mapas do estado de São Paulo da Figura 5 ilustram a frequência de sinistros das outras coberturas (assistência 24 horas etc.) nos diversos municípios, onde as cores indicam o nível de risco (segundo critérios estabelecidos) e o tamanho das bolhas é proporcional ao valor da frequência. É possível notar que há regiões no estado de São Paulo com maior concentração de municípios de alto risco enquanto há regiões majoritariamente de baixo risco, direcionando o uso de regiões de circulação para construir os grupos de risco das outras coberturas.

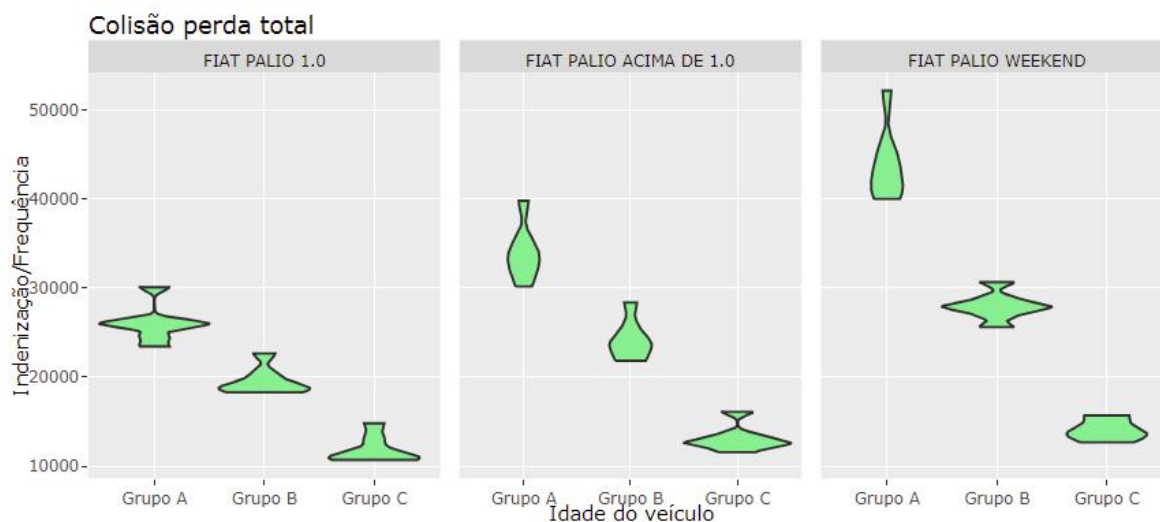


Figura 6 – Distribuição das indenizações de colisão com perda total do *Fiat Palio* ponderada por unidades expostas em função do subgrupo do modelo e grupo de idade do veículo (seleção Brasil e todos os anos de modelo).

A distribuição do valor das indenizações da Figura 6 ilustra indícios de assimetria positiva para o valor dos sinistros de perda total do *Fiat Palio*, dado que um sinistro ocorreu. Além disso, traz evidências que o grupo de idade do veículo e o subgrupo do modelo estão claramente associados aos valores das indenizações por perda total, visto que se relacionam ao valor dos veículos. Desta maneira, há indícios que justificam o uso de distribuições de probabilidade assimétricas positiva para a modelagem da severidade de colisões por perda total, além da inclusão do grupo de idade do veículo e do subgrupo do modelo para os grupos de risco dessa cobertura.

APÊNDICE B – Codificação dos Subgrupos dos Modelos de Veículos

- **Chevrolet Celta:**
 - GM CHEVROLET CELTA 1.0: código 1
 - GM CHEVROLET CELTA ACIMA DE 1.0: código 1
- **Chevrolet Corsa:**
 - GM CHEVROLET CORSA 1.0: código 1
 - GM CHEVROLET CORSA ACIMA DE 1.0: código 2
- **Chevrolet Onix:**
 - GM CHEVROLET ONIX: código 1
- **Chevrolet Prisma:**
 - GM CHEVROLET PRISMA: código 1
- **Fiat Argo:**
 - FIAT ARGO: código 1
- **Fiat Palio:**
 - FIAT PALIO 1.0: código 1
 - FIAT PALIO ACIMA DE 1.0: código 2
 - FIAT PALIO WEEKEND: código 3
- **Fiat Siena:**
 - Fiat Grand Siena: código 1
 - FIAT SIENA 1.0: código 2
 - FIAT SIENA ACIMA DE 1.0: código 3
- **Fiat Strada:**
 - FIAT STRADA: código 1
- **Fiat Uno:**
 - FIAT UNO 1.0: código 1
 - FIAT UNO ACIMA DE 1.0: código 2

- Fiat Uno Vivace: código 3
- Fiat Uno Way: código 3
- **Ford Fiesta:**
 - FORD FIESTA 1.0: código 1
 - FORD FIESTA ACIMA DE 1.0: código 2
- **Ford Ka:**
 - FORD KA 1.0: código 1
 - FORD KA ACIMA DE 1.0: código 2
- **Hyundai HB20:**
 - HYUNDAI HB20: código 1
- **Jeep Compass:**
 - JEEP-COMPASS: código 1
- **Renault Kwid:**
 - RENAULT KWID: código 1
- **Volkswagen Fox:**
 - VW VOLKSWAGEN CROSSFOX: código 1
 - VW VOLKSWAGEN FOX 1.0: código 2
 - VW VOLKSWAGEN FOX ACIMA DE 1.0: código 3
 - VW VOLKSWAGEN SPACEFOX: código 4
- **Volkswagen Gol:**
 - VW VOLKSWAGEN GOL 1.0: código 1
 - VW VOLKSWAGEN GOL ACIMA DE 1.0: código 2
- **Volkswagen Polo:**
 - VW VOLKSWAGEN POLO: código 1

APÊNDICE C – Resultados de Prêmios e Indenizações para o Ano de 2019 por Cobertura

Tabela 11 – Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, detalhados por cobertura.

Cobertura/Modelo	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Roubo/Furto + Incêndio			
Chevrolet Celta	37.380.368,62	27.131.684,00	0,73
Chevrolet Corsa	35.770.814,96	26.805.063,00	0,75
Chevrolet Onix	149.321.162,15	104.191.370,00	0,70
Chevrolet Prisma	69.181.638,19	56.807.108,00	0,82
Fiat Argo	16.294.395,68	9.808.629,00	0,60
Fiat Palio	125.415.126,94	100.950.334,00	0,80
Fiat Siena	80.163.150,86	76.011.915,00	0,95
Fiat Strada	84.227.688,62	58.645.071,00	0,70
Fiat Uno	72.986.588,57	55.015.843,00	0,75
Ford Fiesta	87.515.151,21	62.016.449,00	0,71
Ford Ka	59.324.816,81	54.180.850,00	0,91
Hyundai HB20	245.324.389,17	139.803.920,00	0,57
Jeep Compass	33.276.227,95	24.475.519,00	0,74
Renault Kwid	6.618.141,28	6.025.529,00	0,91
Volkswagen Fox	155.261.432,69	122.829.307,00	0,79
Volkswagen Gol	120.383.724,52	82.809.596,00	0,69
Volkswagen Polo	33.072.342,10	18.091.383,00	0,55
Subtotal	1.411.517.160,32	1.025.599.570,00	0,73
Colisão Parcial			
Chevrolet Celta	39.684.813,29	24.994.996,00	0,63
Chevrolet Corsa	59.156.412,03	39.025.814,00	0,66
Chevrolet Onix	141.655.280,81	133.199.024,00	0,94
Chevrolet Prisma	70.664.661,01	74.094.250,00	1,05
Fiat Argo	20.611.962,14	18.506.553,00	0,90
Fiat Palio	102.283.125,17	81.637.917,00	0,80

(continua na próxima página)

Tabela 11 – (continuação)

Cobertura/Modelo	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Fiat Siena	73.345.544,34	52.999.314,00	0,72
Fiat Strada	63.438.605,10	58.431.140,00	0,92
Fiat Uno	59.705.897,67	43.242.694,00	0,72
Ford Fiesta	106.495.500,12	86.670.243,00	0,81
Ford Ka	70.804.647,56	79.456.170,00	1,12
Hyundai HB20	182.242.331,01	147.233.876,00	0,81
Jeep Compass	39.562.146,06	41.681.594,00	1,05
Renault Kwid	21.645.293,68	20.147.620,00	0,93
Volkswagen Fox	134.451.980,34	100.360.207,00	0,75
Volkswagen Gol	114.873.845,87	91.861.229,00	0,80
Volkswagen Polo	31.949.095,95	31.349.151,00	0,98
Subtotal	1.332.571.142,15	1.124.891.792,00	0,84
Colisão Perda Total			
Chevrolet Celta	22.685.393,09	27.695.557,00	1,22
Chevrolet Corsa	33.157.216,60	45.727.703,00	1,38
Chevrolet Onix	91.301.235,98	79.934.236,00	0,88
Chevrolet Prisma	54.642.953,84	52.993.746,00	0,97
Fiat Argo	9.806.340,84	8.918.444,00	0,91
Fiat Palio	64.980.825,06	77.273.242,00	1,19
Fiat Siena	36.678.602,02	50.312.819,00	1,37
Fiat Strada	54.677.645,53	57.315.058,00	1,05
Fiat Uno	45.701.985,67	46.241.017,00	1,01
Ford Fiesta	69.830.085,94	78.407.099,00	1,12
Ford Ka	57.582.633,37	52.274.591,00	0,91
Hyundai HB20	120.857.579,60	95.393.049,00	0,79
Jeep Compass	32.964.595,27	28.887.865,00	0,88
Renault Kwid	13.207.384,57	12.572.096,00	0,95
Volkswagen Fox	56.656.197,79	69.047.743,00	1,22
Volkswagen Gol	68.761.259,72	72.770.554,00	1,06
Volkswagen Polo	22.000.168,25	23.028.619,00	1,05
Subtotal	855.492.103,14	878.793.438,00	1,03
Outras Coberturas			
Chevrolet Celta	18.097.864,15	22.770.548,00	1,26
Chevrolet Corsa	28.032.779,88	33.302.755,00	1,19
Chevrolet Onix	45.875.427,51	48.865.554,00	1,07

(continua na próxima página)

Tabela 11 – (continuação)

Cobertura/Modelo	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Chevrolet Prisma	22.796.393,60	31.406.453,00	1,38
Fiat Argo	4.691.276,72	4.457.926,00	0,95
Fiat Palio	39.003.128,24	46.666.449,00	1,20
Fiat Siena	25.038.443,46	29.745.609,00	1,19
Fiat Strada	13.452.354,72	20.675.811,00	1,54
Fiat Uno	26.663.521,70	31.972.972,00	1,20
Ford Fiesta	40.289.740,16	52.799.483,00	1,31
Ford Ka	36.510.866,33	43.378.878,00	1,19
Hyundai HB20	58.903.683,30	73.706.337,00	1,25
Jeep Compass	17.204.876,52	20.422.608,00	1,19
Renault Kwid	5.265.325,73	4.967.187,00	0,94
Volkswagen Fox	45.808.228,47	49.938.170,00	1,09
Volkswagen Gol	35.419.447,59	43.330.983,00	1,22
Volkswagen Polo	15.539.273,04	13.977.879,00	0,90
Subtotal	478.592.631,12	572.385.602,00	1,20
Total	4.078.173.036,73	3.601.670.402,00	0,88

APÊNDICE D – Resultados de Prêmios e Indenizações para o Ano de 2019 por Modelo de Veículo

Tabela 12 – Resultados de prêmios estimados, indenizações e sinistralidade para o ano de 2019, detalhados por modelo de veículo.

Cobertura/Modelo	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Chevrolet Celta			
Roubo/Furto e Incêndio	37.380.368,62	27.131.684,00	0,73
Colisão Parcial	39.684.813,29	24.994.996,00	0,63
Colisão Perda Total	22.685.393,09	27.695.557,00	1,22
Outras coberturas	18.097.864,15	22.770.548,00	1,26
Subtotal	117.848.439,15	102.592.785,00	0,87
Chevrolet Corsa			
Roubo/Furto e Incêndio	35.770.814,96	26.805.063,00	0,75
Colisão Parcial	59.156.412,03	39.025.814,00	0,66
Colisão Perda Total	33.157.216,60	45.727.703,00	1,38
Outras coberturas	28.032.779,88	33.302.755,00	1,19
Subtotal	156.117.223,47	144.861.335,00	0,93
Chevrolet Onix			
Roubo/Furto e Incêndio	149.321.162,15	104.191.370,00	0,70
Colisão Parcial	141.655.280,81	133.199.024,00	0,94
Colisão Perda Total	91.301.235,98	79.934.236,00	0,88
Outras coberturas	45.875.427,51	48.865.554,00	1,07
Subtotal	428.153.106,45	366.190.184,00	0,86
Chevrolet Prisma			
Roubo/Furto e Incêndio	69.181.638,19	56.807.108,00	0,82
Colisão Parcial	70.664.661,01	74.094.250,00	1,05
Colisão Perda Total	54.642.953,84	52.993.746,00	0,97
Outras coberturas	22.796.393,60	31.406.453,00	1,38
Subtotal	217.285.646,64	215.301.557,00	0,99
Fiat Argo			

(continua na próxima página)

Tabela 12 – (continuação)

Modelo/Cobertura	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Roubo/Furto e Incêndio	16.294.395,68	9.808.629,00	0,60
Colisão Parcial	20.611.962,14	18.506.553,00	0,90
Colisão Perda Total	9.806.340,84	8.918.444,00	0,91
Outras coberturas	4.691.276,72	4.457.926,00	0,95
Subtotal	51.403.975,38	41.691.552,00	0,81
Fiat Palio			
Roubo/Furto e Incêndio	125.415.126,94	100.950.334,00	0,80
Colisão Parcial	102.283.125,17	81.637.917,00	0,80
Colisão Perda Total	64.980.825,06	77.273.242,00	1,19
Outras coberturas	39.003.128,24	46.666.449,00	1,20
Subtotal	331.682.205,41	306.527.942,00	0,92
Fiat Siena			
Roubo/Furto e Incêndio	80.163.150,86	76.011.915,00	0,95
Colisão Parcial	73.345.544,34	52.999.314,00	0,72
Colisão Perda Total	36.678.602,02	50.312.819,00	1,37
Outras coberturas	25.038.443,46	29.745.609,00	1,19
Subtotal	215.225.740,68	209.069.657,00	0,97
Fiat Strada			
Roubo/Furto e Incêndio	84.227.688,62	58.645.071,00	0,70
Colisão Parcial	63.438.605,10	58.431.140,00	0,92
Colisão Perda Total	54.677.645,53	57.315.058,00	1,05
Outras coberturas	13.452.354,72	20.675.811,00	1,54
Subtotal	215.796.293,97	195.067.080,00	0,90
Fiat Uno			
Roubo/Furto e Incêndio	72.986.588,57	55.015.843,00	0,75
Colisão Parcial	59.705.897,67	43.242.694,00	0,72
Colisão Perda Total	45.701.985,67	46.241.017,00	1,01
Outras coberturas	26.663.521,70	31.972.972,00	1,20
Subtotal	205.057.993,61	176.472.526,00	0,86
Ford Fiesta			
Roubo/Furto e Incêndio	87.515.151,21	62.016.449,00	0,71
Colisão Parcial	106.495.500,12	86.670.243,00	0,81
Colisão Perda Total	69.830.085,94	78.407.099,00	1,12
Outras coberturas	40.289.740,16	52.799.483,00	1,31

(continua na próxima página)

Tabela 12 – (continuação)

Modelo/Cobertura	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Subtotal	304.130.477,43	279.893.274,00	0,92
Ford Ka			
Roubo/Furto e Incêndio	59.324.816,81	54.180.850,00	0,91
Colisão Parcial	70.804.647,56	79.456.170,00	1,12
Colisão Perda Total	57.582.633,37	52.274.591,00	0,91
Outras coberturas	36.510.866,33	43.378.878,00	1,19
Subtotal	224.222.964,07	229.290.489,00	1,02
Hyundai HB20			
Roubo/Furto e Incêndio	245.324.389,17	139.803.920,00	0,57
Colisão Parcial	182.242.331,01	147.233.876,00	0,81
Colisão Perda Total	120.857.579,60	95.393.049,00	0,79
Outras coberturas	58.903.683,30	73.706.337,00	1,25
Subtotal	607.327.983,08	456.137.182,00	0,75
Jeep Compass			
Roubo/Furto e Incêndio	33.276.227,95	24.475.519,00	0,74
Colisão Parcial	39.562.146,06	41.681.594,00	1,05
Colisão Perda Total	32.964.595,27	28.887.865,00	0,88
Outras coberturas	17.204.876,52	20.422.608,00	1,19
Subtotal	123.007.845,80	115.467.586,00	0,94
Renault Kwid			
Roubo/Furto e Incêndio	6.618.141,28	6.025.529,00	0,91
Colisão Parcial	21.645.293,68	20.147.620,00	0,93
Colisão Perda Total	13.207.384,57	12.572.096,00	0,95
Outras coberturas	5.265.325,73	4.967.187,00	0,94
Subtotal	46.736.145,26	43.712.432,00	0,94
Volkswagen Fox			
Roubo/Furto e Incêndio	155.261.432,69	122.829.307,00	0,79
Colisão Parcial	134.451.980,34	100.360.207,00	0,75
Colisão Perda Total	56.656.197,79	69.047.743,00	1,22
Outras coberturas	45.808.228,47	49.938.170,00	1,09
Subtotal	392.177.839,29	342.175.427,00	0,87
Volkswagen Gol			
Roubo/Furto e Incêndio	120.383.724,52	82.809.596,00	0,69
Colisão Parcial	114.873.845,87	91.861.229,00	0,80

(continua na próxima página)

Tabela 12 – (continuação)

Modelo/Cobertura	Prêmio (R\$)	Indenização (R\$)	Sinistralidade
Colisão Perda Total	68.761.259,72	72.770.554,00	1,06
Outras coberturas	35.419.447,59	43.330.983,00	1,22
Subtotal	339.438.277,70	290.772.362,00	0,86
Volkswagen Polo			
Roubo/Furto e Incêndio	33.072.342,10	18.091.383,00	0,55
Colisão Parcial	31.949.095,95	31.349.151,00	0,98
Colisão Perda Total	22.000.168,25	23.028.619,00	1,05
Outras coberturas	15.539.273,04	13.977.879,00	0,90
Subtotal	102.560.879,34	86.447.032,00	0,84
Total	4.078.173.036,73	3.601.670.402,00	0,88