

# CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ  
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v. 58, p. 01-27, 2025  
DOI: 10.12957/cadest.2025.94260

## PROTEÇÃO PATRIMONIAL MUTUALISTA: COMPARAÇÃO DE CUSTEIO COM O SEGURO AUTOMÓVEL TRADICIONAL

Eduardo Fraga L. de Melo  
Instituto de Matemática e Estatística – IME / UERJ  
Escola de Matemática Aplicada – EMAp / FGV  
[eduardofraga@ime.uerj.br](mailto:eduardofraga@ime.uerj.br)

Maria Tereza das Graças  
Universidade Federal do Rio de Janeiro / UFRJ  
[mt.tereza@hotmail.com](mailto:mt.tereza@hotmail.com)

Paulo Andrés Toro  
Universidade Federal do Rio de Janeiro / UFRJ  
[pauloat@gmail.com](mailto:pauloat@gmail.com)

### Resumo

*O setor de seguros vem passando por transformações marcantes nos últimos tempos. Uma destas transformações é a introdução da chamada proteção patrimonial mutualista, regulamentada pela recém-publicada Lei Complementar n. 213/2025. Em comparação com o seguro “tradicional”, a proteção mutualista possui diferenças, em especial, no que se refere ao custeio. Em mercados internacionais, essa operação é chamada de seguros descentralizados ou mutual-aid. Ao contrário do seguro tradicional, onde o prêmio é estabelecido ex-ante, nesta operação, o “prêmio” é definido de forma expost por meio do rateio das despesas da carteira com sinistros. Com o uso de modelo de machine learning, que se mostrou o melhor em performance, este estudo compara os custos do seguro automóvel tradicional usando dados reais e públicos de apólices subscritas por seguradoras entre os anos de 2012 e 2020 (último ano disponível dos dados) com aquele que seria cobrado em um arranjo mutualista. Calculamos também o carregamento máximo sobre o prêmio que as administradoras de proteção patrimonial mutualista poderiam ter cobrado nestas apólices para que fossem, em média, tão atrativas para os segurados quanto o seguro tradicional. A presente pesquisa vem em momento crucial para o mercado de seguros e busca endereçar o custeio destes dois tipos de operação.*

**Palavras-chave:** Proteção Patrimonial Mutualista; Seguro Descentralizado; Seguro Automóvel; Custeio.

## 1. Introdução

O setor de seguros desempenha um papel fundamental na mitigação de riscos e na proteção financeira de indivíduos e empresas. Tradicionalmente, o seguro é operado em um modelo centralizado, no qual seguradoras gerenciam a precificação, emissão de apólices, recepção de prêmios e posterior liquidação de sinistros. No entanto, com o avanço da economia compartilhada, das finanças descentralizadas, tecnologias de blockchain e contratos inteligentes, surgiu no mundo um novo paradigma: os seguros descentralizados.

No Brasil, essa operação não estava prevista em Lei até recentemente. Apesar de não previsto, foi observado, ao longo do tempo, o aparecimento de associações de proteção operando várias linhas de negócio (especialmente automóvel). Há alegações de que os preços praticados por estas associações seriam mais baixos que aqueles dos seguros congêneres oferecidos pelas seguradoras dentro do mercado regulado ou no modelo centralizado. Atualmente, não é possível comparar, em estudo científico, os preços destas duas modalidades. Não há registros formais dos preços ao cliente final praticados pelas associações ao longo do tempo, afinal não participavam do mercado regulado. Neste ambiente, além de não terem os custos regulatórios, havia também um vácuo tributário e os consumidores não estavam devidamente protegidos pelo arcabouço legal de seguros.

A recente publicação da Lei Complementar 213, de 15 de janeiro de 2025 (BRASIL, 2025), acabou possibilitando a regular operação destas coberturas no país, por meio do modelo de proteção patrimonial mutualista, para quase todas as linhas de negócio securitárias. Isso introduziu mudanças significativas no Sistema Nacional de Seguros Privados no Brasil, como:

- a) operações de proteção patrimonial mutualista: a lei regulamenta essas operações, permitindo que grupos de pessoas se associem para compartilhar riscos predeterminados por meio de rateio de despesas. Essas operações devem ser geridas por administradoras especializadas, que estarão sob supervisão da Superintendência de Seguros Privados (SUSEP);
- b) ampliação de atuação das Sociedades Cooperativas de seguros: anteriormente restritas a ramos específicos, como seguros agrícolas, de saúde e de acidentes de trabalho, as cooperativas de seguros agora estão autorizadas a operar em qualquer

ramo de seguros privados, exceto em contratos estruturados em capitalização e repartição de capitais de cobertura;

- c) regularização de associações existentes: a lei estabelece regras e condições para a regularização de associações que, na data de sua publicação, já operavam no mercado sem uma previsão legal específica. Essas associações deverão se adequar às novas disposições para continuar suas atividades.

Em virtude do grande interesse no efeito desta Lei nos preços de seguros, especificamente no seguro automóvel, neste artigo, compararemos, de forma empírica, qual seria o custo de uma proteção de seguro automóvel no modelo tradicional e na proteção patrimonial mutualista (descentralizado) com base em dados reais e públicos para os anos de 2012 a 2020. Além disso, calculamos também o carregamento máximo sobre o prêmio que as administradoras de proteção patrimonial mutualista poderiam ter cobrado nestas apólices para serem mais atrativas para os segurados que o seguro tradicional centralizado.

Desta forma, a presente pesquisa vem em momento crucial para o mercado de seguros e busca endereçar a comparação no custeio destes dois tipos de operação para uma linha de negócio de grande relevância.

### **1.1 Custeio da proteção patrimonial mutualista (seguro descentralizado)**

A proteção patrimonial mutualista, ou os seguros descentralizados, também conhecidos como *DeFi insurance* ou *peer-to-peer (P2P)*, *mutual aid*, *collaborative insurance* ou *crowdinsurance*, costumam utilizar redes distribuídas para automatizar e tornar mais transparentes os processos de subscrição e pagamento de sinistros, reduzindo custos operacionais e eliminando intermediários. Seu conceito remonta às raízes mais antigas do seguro, onde os indivíduos se organizam em grupos e o mecanismo de compensação entre os membros consiste no uso das contribuições de muitos, por livres iniciativa e vontade para obter cobertura, para equilibrar os infortúnios de poucos.

Antes de comentar, de forma introdutória, sobre aspectos atuariais que distinguem os seguros descentralizados dos “centralizados” e “híbridos”, cabe diferenciá-los, remetendo a Feng (2023) e Dhaene (2024).

#### **1.1.1 Seguros “centralizados”**

As características principais dos seguros centralizados (tradicionais) incluem:

- a. há transferência de risco para um ente central (por ex. seguradora);

- b. cada segurado na carteira é compensado *ex-post* pela seguradora por sua perda;
- c. em retorno, a seguradora cobra *ex-ante* um prêmio de cada segurado e constitui seu capital de solvência;
- d. prêmios e capital são estabelecidos de forma que a probabilidade de sua soma exceder as perdas da carteira seja suficientemente alta;
- e. o custeio segue um princípio de cálculo do prêmio (aqueles definidos em Teoria do Risco em Kaas *et al.*, (2008) - ex: valor esperado das perdas, valor esperado com carregamento de segurança, princípio do percentil, etc.

### 1.1.2 Seguros descentralizados (“mutual aid”)

As características principais dos seguros descentralizados (proteção mutualista) incluem:

- a. mecanismo de compartilhamento de risco, onde os participantes de um grupo compartilham os riscos entre eles;
- b. cada participante do grupo é compensado *ex-post* pelo grupo por sua perda;
- c. em retorno, cada participante paga *ex-post* uma contribuição para o grupo;
- d. a abordagem descentralizada é totalmente autofinanciada;
- e. as contribuições seguem uma regra de compartilhamento de risco, estabelecida previamente e de tal forma que a soma das contribuições dos participantes do grupo é igual à perda total do grupo.

Para qualquer regra de compartilhamento de risco, o fato de a abordagem ser completamente autofinanciada significa, matematicamente, que (1):

$$\sum_{i=1}^n P_i = S = \sum_{i=1}^n X_i \quad (1)$$

onde:  $P_i$  é o prêmio puro do membro individual  $i$ ;  $S$  é a perda total observada da carteira e  $X_i$  é a perda do membro individual  $i$ .

Como se pode observar, o prêmio de fato que o membro paga em um arranjo de proteção patrimonial mutualista é aleatório, pois depende diretamente da perda total da carteira ainda a ser observada.

### 1.1.3 Seguros descentralizados “híbridos”

São aqueles em que há um componente característico de compartilhamento de risco e há a transferência de um excedente para uma seguradora. Ou seja, até um determinado limite de perdas, o risco é compartilhado entre os membros do grupo. Caso as perdas superem o limite, o excedente é de responsabilidade de uma seguradora. Neste

tipo de modelo, é possível a cobrança *ex-ante*, com posterior *cashback* ou devolução, caso as perdas fiquem abaixo do limite citado.

## 1.2 Regras de compartilhamento de riscos (RC)

As regras de compartilhamento (RC) dos riscos definem qual será a contribuição de cada membro para a cobertura das perdas do grupo. Alguns exemplos de RC (presentes em DHAENE, 2024):

- i) Uniforme: regra mais trivial, perda compartilhada de forma igual para todos do grupo.
- ii) Baseada em algum uso do bem segurado: ex. compartilhamento segue quilometragem utilizada do veículo.
- iii) Uniforme baseada em sub-grupos: grupo dividido em sub-grupos por alguma característica (idade, por ex.) e, dentro desses sub-grupos, o compartilhamento é uniforme.
- iv) Média proporcional: cada participante paga um percentual da perda total definido de acordo com a expectativa de perda que ele traz para o grupo. Critério mais observado na prática, conforme comenta Dhaene (2024) (2):

$$P_i = S \cdot \frac{E(X_i)}{E(S)} \quad (2)$$

onde  $X_i$  é a perda do membro individual  $i$ ,  $S$  é a perda total observada da carteira e  $E(S) = \sum_i E(X_i)$ .

Ex.: grupo hipotético com apenas dois membros. O membro A possui carro mais caro, utiliza mais ou possui maior perfil de risco, assim tem valor esperado de perdas igual a 75 ( $E(X_A) = 75$ ). O membro B possui carro mais barato, utiliza menos ou possui menor perfil de risco, assim tem valor esperado de perdas igual a 25 ( $E(X_B) = 25$ ). Os percentuais de compartilhamento serão 75% e 25% nesse grupo. Se as perdas do grupo forem 80 ( $S = 80$ ), as contribuições serão 60 ( $= 80 * \frac{75}{(75+25)}$ ) e 20 ( $= 80 * \frac{25}{(75+25)}$ ), respectivamente, para A e B. Se as perdas do grupo forem 200, serão 150 ( $= 200 * \frac{75}{(75+25)}$ ) e 50 ( $= 200 * \frac{25}{(75+25)}$ ).

- v) Média condicional: cada participante paga o valor esperado das perdas que ele traz para o grupo condicionado à perda agregada do grupo (3):

$$P_i = E(X_i|S) \quad (3)$$

A regra de compartilhamento 5 (média condicional) depende da especificação de distribuições de probabilidade das perdas individuais ( $X_i$ ) de cada membro do grupo mutualista. Isso se torna um obstáculo para seu uso caso sejam aplicados algoritmos não probabilísticos para o cálculo do rateio. Por outro lado, essa regra cumpre todas as propriedades descritas em Dhaene (2024). Essas propriedades são relevantes para uma regra de compartilhamento.

A regra de compartilhamento 4 (média proporcional) cumpre a maioria das propriedades descritas em Dhaene (2024) e, conforme informações naquele trabalho, é a regra mais utilizada na prática. Para essa regra de compartilhamento, o valor esperado do prêmio de um membro do grupo é igual ao valor esperado das perdas desse membro individualmente. Vejamos:

Propriedade da “Adequação Atuarial”. Uma regra de compartilhamento de riscos em um seguro descentralizado terá a propriedade da adequação atuarial se e somente se (4):

$$E(P_i) = E(X_i) \quad (4)$$

onde  $P_i$  é o prêmio puro (sem qualquer carregamento) do seguro descentralizado para o indivíduo  $i$  e  $X_i$  é a variável aleatória que representa a perda do indivíduo  $i$  com seu patrimônio.

Corolário 1. A regra de compartilhamento de Média proporcional cumpre a propriedade de adequação atuarial, onde o valor esperado do prêmio de um membro do grupo é igual ao valor esperado das perdas desse membro.

Demonstração (5):

$$P_i = S \cdot \frac{E(X_i)}{E(S)} \quad (5)$$

onde:  $X_i$  é a perda de um membro individual;  $S$  é a perda total da carteira.

Tomando a esperança em ambos os lados da equação (5), temos (6):

$$E(P_i) = E \left[ S \cdot \frac{E(X_i)}{E(S)} \right] = \frac{E(X_i)}{E(S)} E[S] = E(X_i) \quad (6)$$

Observa-se, por exemplo, que a regra de compartilhamento uniforme não cumpre com esta propriedade. Afinal (7):

$$P_i = \frac{S}{n} \quad (7)$$

onde  $n$  é o número de membros na carteira (ou seja, é distribuição uniforme das perdas da carteira entre os membros do grupo). Aplicando o operador de esperança nos dois lados da equação (3), temos (8):

$$E(P_i) = E\left[\frac{S}{n}\right] = \frac{1}{n}E(S) \quad (8)$$

$\frac{1}{n}E(S)$  somente é igual a  $E(X_i)$  se e somente se  $X_i$  possuírem médias equivalentes. Portanto, não se pode afirmar que essa regra de compartilhamento possui a propriedade da adequação atuarial.

Outro ponto que merece atenção: o valor esperado do prêmio ser equivalente ao valor esperado das perdas não significa afirmar que a contribuição do membro será igual ao valor esperado de suas perdas. Afinal, o valor absoluto da contribuição é definido *ex-post*, então é algo incerto no momento inicial.

No seguro descentralizado há uma troca da incerteza que diz respeito à ocorrência de um sinistro, em que a pessoa se responsabilizaria por si mesma, por uma incerteza menor pelo agrupamento de riscos (prêmio *ex-post*). Em outras palavras, é a troca de uma despesa com incerteza maior por uma despesa com incerteza menor.

No seguro centralizado há a troca da incerteza que diz respeito à ocorrência de um sinistro, em que a pessoa se responsabilizaria por si mesma, por um valor certo (prêmio *ex-ante*). Em outras palavras, é a troca de uma despesa com incerteza por uma despesa sem incerteza.

Em ambos os casos, essas trocas se justificam economicamente (e, por isso, existe o mercado de seguros) em um ambiente tão somente com indivíduos com função utilidade crescente (preferência por mais do que menos, financeiramente) e aversão ao risco. A demonstração pode ser consultada no capítulo 1 de Ferreira (2002).

Em resumo, neste estudo:

- a) modelamos as perdas de seguro automóvel com dados reais e públicos utilizando algoritmo de machine learning (*XGBoost*, que foi o de melhor performance) de forma a calcular qual seria o prêmio de cada apólice em uma operação de proteção mutualista (seguro descentralizado), utilizando a regra de compartilhamento de risco da média proporcional (a mais utilizada na prática);
- b) calculamos a diferença entre o prêmio do seguro tradicional (informado no banco de dados) e qual seria o prêmio puro da proteção, considerando que todos os automóveis da amostra, em determinado ano, estão na carteira. Realizamos a

análise dessa diferença em função da Importância Segurada das apólices, da faixa etária do condutor e em função da região geográfica do automóvel a fim de identificar se há alguma relação entre a diferença dos prêmios por alguma dessas variáveis/características;

- c) o banco de dados possui a informação de prêmio de seguro por apólice bruto de carregamentos de segurança, custo de aquisição, despesas administrativas e margem de lucro. Já o prêmio do seguro descentralizado é puro. Dessa forma, calculamos qual seria o carregamento que faz com que a média dos prêmios da proteção (seguro descentralizado) equivalha à média dos prêmios do seguro tradicional.

Todos os modelos e cálculos foram realizados no RStudio (pacote R 4.3.0) com as bibliotecas *xgboost*, *randomForest*, *bst* e *gamlss*.

O artigo está organizado da seguinte forma: na seção 2 é realizada uma revisão de literatura sobre os temas explorados neste artigo; na seção 3 são apresentados os dados utilizados (que são públicos). Como o estudo tem métodos e resultados sequenciais, como explicitado no resumo logo acima, entendemos que o leitor acompanhará melhor o desenvolvimento com a apresentação dos métodos seguidos por seus resultados, de forma conjunta e sequencial na seção 4 (metodologia e resultados). Por fim, são feitas considerações finais na seção 5.

## **2. Revisão de Literatura**

Grande parte da literatura sobre seguros concentra-se nos modelos centralizados, nos quais seguradoras atuam como operadores dos riscos subscritos e indenização das perdas. No entanto, o avanço das tecnologias descentralizadas tem impulsionado novas abordagens no setor, particularmente com o uso de blockchain e contratos inteligentes (BUTERIN, 2014).

### **2.1 Seguros Centralizados**

Os modelos de seguros centralizados são amplamente estudados na literatura devido à sua relevância econômica e sua estrutura regulatória consolidada. Sobre este último, as normas afetas à solvência de seguros de forma geral, e sobre produtos de seguro de automóveis, em particular, são CNSP (2021) e SUSEP (2021), respectivamente. As seguradoras desempenham um papel crucial na precificação de riscos, na criação de reservas financeiras para sinistros e na proteção contra fraudes (KUNREUTHER *et al.*,

2013). Para práticas no Brasil, sugerimos Ferreira (2002) e Ferreira e Mano (2009). No entanto, pesquisas também apontam desafios, como custos administrativos elevados, processos burocráticos e assimetrias de informação que podem prejudicar a experiência dos segurados (AKERLOF, 2017).

O mercado de seguro automóvel, especificamente, tem sido alvo de estudos sobre precificação baseada em big data e inteligência artificial para mitigar problemas de seleção adversa e risco moral (KITCHENS, 2019). Apesar dos avanços tecnológicos, a centralização ainda gera preocupações com transparência e eficiência na liquidação de sinistros (GRACE, 2025).

## **2.2 Seguros Descentralizados (proteção patrimonial mutualista)**

O surgimento dos seguros descentralizados também está associado ao crescimento das finanças descentralizadas (DeFi) e à adoção de blockchain como uma tecnologia confiável para garantir transparência e automação (BORSELLI, 2019). Os contratos inteligentes, inicialmente propostos por Szabo (1997), permitem que apólices de seguro sejam executadas automaticamente quando as condições predefinidas são atendidas, eliminando intermediários e reduzindo custos operacionais (BUTERIN, 2014).

Estudos recentes em mercados internacionais destacam vantagens dos seguros descentralizados, como maior acessibilidade global e menor vulnerabilidade a fraudes e manipulações (ARNOTT & STIGLITZ, 1988). No entanto, desafios como a dependência de fontes externas para validar eventos e obstáculos regulatórios ainda são questões em aberto na literatura (SCHÄR, 2021). Cabe ressaltar que seguros descentralizados fora do Brasil, e sua pesquisa, estão intimamente relacionados com o uso de tecnologias descentralizadas.

## **2.3 Aplicação no Seguro Automóvel**

A aplicação de seguros descentralizados no setor automotivo tem sido estudada sob diferentes perspectivas. De acordo com Werbach (2018), a adoção de contratos inteligentes em seguros veiculares pode reduzir significativamente o tempo de liquidação de sinistros e aumentar a transparência para segurados e seguradoras. Além disso, alguns modelos propõem o uso de dispositivos IoT para monitoramento em tempo real e precificação dinâmica com base no comportamento dos motoristas (RUBINSTEIN CAVALCANTI, 2018).

Apesar das inovações, a adoção dos seguros descentralizados no seguro automóvel ainda enfrenta desafios regulatórios e de aceitação pelo mercado. Como apontado por Zetzsche *et al.* (2020), a transição para modelos descentralizados depende não apenas de avanços tecnológicos, mas também de marcos regulatórios que garantam a segurança dos consumidores. Enquanto os seguros centralizados oferecem previsibilidade e proteção regulatória, os descentralizados propõem maior eficiência e transparência. O presente estudo busca contribuir para esse debate, analisando as implicações práticas da aplicação desses modelos no seguro automóvel.

#### 2.4 Precificação de seguros com machine learning

Seja  $(Z_1, D_1, Y_1), \dots, (Z_n, D_n, Y_n)$  uma amostra de tripletos independentes e identicamente distribuídos, em que, para a  $i$ -ésima unidade amostral,  $Z_i \in R^p$  é um vetor de preditores,  $D_i \in N$  denota a frequência de sinistros ocorridos para a  $i$ -ésima unidade amostral e  $Y_i \in R^+$  é a severidade do sinistro da  $i$ -ésima unidade amostral. Portanto, a perda esperada por unidade amostral (membro)  $i$  é:  $E(X_i) = E(D_i)E(Y_i)$ .

A modelagem preditiva neste contexto é implementada usando dois modelos aplicados em sucessão: um primeiro modelo prevê a frequência de sinistros a partir dos preditores disponíveis, enquanto um segundo modelo prevê a severidade média dos sinistros a partir dos preditores disponíveis. Por fim, a perda predita para o membro  $i$  é o produto dos dois valores preditos anteriormente.

Regressões costumam ser escolhas paramétricas para os estágios de frequência e severidade, respectivamente. Recentemente, algoritmos modernos de aprendizado de máquina, como florestas aleatórias (BREIMAN, 2001), *boosting*, *XGBoost*, foram aplicados com sucesso neste cenário, com ganhos substanciais de desempenho preditivo (HENCKAERTS *et al.*, 2021 e SU & BAI, 2020).

### 3. Dados

Os dados são provenientes da base de dados AUTOSEG<sup>1</sup> entre os anos 2012 e 2020 (último ano disponível). A tabela utilizada para modelagem possui dados de exposição, prêmios, sinistros e importância segurada para a cobertura CASCO, classificados pela chave Categoria Tarifária/Região/Modelo/Ano/Sexo/Faixa Etária<sup>2</sup>.

---

<sup>1</sup> <https://www2.susep.gov.br/menuestatistica/autoseg/principal.aspx>

<sup>2</sup> Tabela "arq\_casco\_comp.csv" de cada base de dados entre os anos 2012 e 2020.

- a. Categoria Tarifária: (i) passeio nacional; (ii) passeio importado; (iii) pick-up; (iv) veículo de carga; (v) motocicleta; (vi) ônibus; (vii) utilitários ou (viii) outros. Utilizamos a categoria passeio nacional neste estudo.
- b. Região: região do veículo, podendo ser 41 regiões do Brasil. Utilizamos as seguintes regiões nesse estudo:
  - 11 – SP/Metropolitana de São Paulo;
  - 13 – SP/Ribeirão Preto e demais municípios de Campinas;
  - 15 – MG/região Sul;
  - 16 – MG/Metropolitana de BH – Centro Oeste – Zona da Mata – C. Vertentes;
  - 18 – RJ/Metropolitana do Rio de Janeiro e
  - 19 – RJ/Interior.
- c. Modelo: modelo do veículo com códigos descritos nas tabelas auxiliares que podem ser obtidas no banco de dados. Utilizamos todos os modelos de veículos dentro da categoria de passeio nacional.
- d. Ano do modelo: ano do modelo do veículo. Como utilizamos os bancos de dados de 2012 a 2020, transformamos essa variável que qualifica o veículo em anos de idade do veículo na data-base do banco. Assim, se o modelo do carro for 2013 no banco de 2012, essa variável recebe o valor 0, e assim sucessivamente, afinal as montadoras lançam no ano  $t$  carros novos do modelo de ano  $t+1$ .
- e. Sexo: (i) masculino; (ii) feminino e (iii) jurídico. Neste estudo, utilizamos os registros com sexos “masculino” e “feminino”.
- f. Faixa etária: (i) não informado; (ii) 18 a 25 anos; (iii) 26 a 35 anos; (iv) 36 a 45 anos; (v) 46 a 55 anos e (vi) maior que 55 anos. Neste estudo, descartamos os registros com faixa etária “não informada”.

Além da chave, essa tabela possui as seguintes variáveis:

- i) IS Média (IS\_MEDIA): a Importância Segurada Média (R\$) representa a média das IS's das apólices incluídas no grupamento, ponderada pela exposição de cada uma delas. Foram retirados registros com IS\_MEDIA menor ou igual a zero.
- ii) Expostos (EXPOSICAO1): 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. Foram retirados registros com EXPOSICAO1 menor ou igual a zero.

- iii) Prêmio médio (PREMIO1): da mesma forma que a IS Média, o prêmio médio (R\$) representa a média dos prêmios das apólices incluídas no grupamento, ponderada pela exposição de cada uma delas.
- iv) FREQ: soma de (a) total do número de sinistros da cobertura roubo/furto (FREQ1); (b) total do número de sinistros da cobertura colisão parcial (FREQ2); (c) total do número de sinistros da cobertura colisão perda total (FREQ3); (d) total do número de sinistros da cobertura incêndio (FREQ4) e (e) total do número de sinistros de outras coberturas, como assistência 24 hs, etc (FREQ9).

Essa varável pode apresentar o valor zero, caso não tenha havido sinistro.

- v) INDENIZ: soma de (a) total de indenizações de sinistros da cobertura roubo/furto (INDENIZ1); (b) total de indenizações de sinistros da cobertura colisão parcial (INDENIZ2); (c) total de indenizações de sinistros da cobertura colisão perda total (INDENIZ3); (d) total de indenizações de sinistros da cobertura incêndio (INDENIZ4) e (e) total de indenizações de sinistros de outras coberturas, como assistência 24 hs, etc (INDENIZ9).

Essa varável pode apresentar o valor zero, caso não tenha havido sinistro, ou valores positivos, por terem ocorridos 1 ou mais sinistros.

Conforme citado acima, restringimos nossas análises aos veículos do tipo passeio nacional, nas regiões SP/Metropolitana de São Paulo; SP/Ribeirão Preto e demais municípios de Campinas; MG/região Sul; MG/Metropolitana de BH – Centro Oeste – Zona da Mata – C. Vertentes; RJ/Metropolitana do Rio de Janeiro e RJ/Interior; com sexos informados masculino ou feminino e com faixa etária informada. Sobre o ano do modelo dos veículos, transformamos essa variável em idade do veículo no ano base. Além disso, agrupamos os veículos com mais de 5 anos de idade na chave para cada um dos anos-bases. Assim, temos 6 níveis de idade de veículo (0, 1, 2, 3, 4 e 5 anos).

Entendemos que esse banco de dados pode trazer informações relevantes comparando as capitais com o interior de estados de SP; MG e RJ para veículos de passeio nacional, que, em teoria, seriam o maior público para a proteção mutualista. Importante

comentar também que essas características presentes no banco de dados foram utilizadas como preditores da indenização nos modelos. Na prática, pode ser que seguradoras e administradoras de proteção patrimonial mutualista utilizem outras características para precificação. Como não temos essa informação, utilizamos todas as que estão disponíveis nestes bancos de dados públicos. Seguem algumas estatísticas de resumo para o banco de dados usado no estudo.

Tabela 1. Índice de Sinistralidade e total de Indenizações, prêmios e exposição por Sexo e por Faixa Etária do condutor para os anos de 2012 a 2020, considerando veículos Passeio Nacional e nas regiões SP/Metropolitana de São Paulo; SP/Ribeirão Preto e demais municípios de Campinas; MG/região Sul; MG/Metropolitana de BH – Centro Oeste – Zona da Mata – C. Vertentes; RJ/Metropolitana do Rio de Janeiro e RJ/Interior.

<b>Sexo</b>	<b>Faixa Etária</b>	<b>Indeniz. (mil R\$)</b>	<b>Prêmio (mil R\$)</b>	<b>Exposição</b>	<b>Sinistralidade</b>
F	Entre 18 e 25 anos	474.754	719.604	534.456	0,66
F	Entre 26 a 35 anos	2.999.269	4.384.966	3.763.071	0,68
F	Entre 36 a 45 anos	3.321.829	4.970.792	4.324.528	0,67
F	Entre 46 a 55 anos	2.769.315	4.207.138	3.675.582	0,66
F	Maior que 55 anos	2.565.874	4.138.918	3.809.728	0,62
M	Entre 18 e 25 anos	734.057	874.310	549.979	0,84
M	Entre 26 a 35 anos	3.538.707	4.446.884	3.342.191	0,80
M	Entre 36 a 45 anos	3.707.037	4.943.246	3.902.581	0,75
M	Entre 46 a 55 anos	3.194.345	4.660.938	3.770.774	0,69
M	Maior que 55 anos	4.325.928	6.719.727	5.783.801	0,64

Fonte: elaborado pelos autores (2025)

A Tabela 1 indica que o banco de dados usado neste estudo totaliza mais de 32 milhões de unidades de exposição ao longo de 9 anos. Uma unidade de exposição equivale a um automóvel exposto por um ano completo. É observado uma sinistralidade maior para condutores do sexo masculino e, para esse sexo, há uma diminuição da sinistralidade à medida que o condutor tem maior idade. A Tabela 2 compara os índices de sinistralidade pelas regiões e idades dos veículos considerados neste artigo. Podemos identificar uma maior sinistralidade nas capitais quando comparadas com o interior para uma mesma unidade da federação (SP, MG ou RJ). Com relação à idade do veículo,

observa-se um aumento da sinistralidade até a idade de 4 anos e partir de então o índice se estabiliza por volta de 68%. Adicionalmente, as exposições por idade do veículo aumentam até 3 anos e depois passam a diminuir. Por esses dois motivos, decidimos agrupar os registros de veículos com mais de 5 anos de idade.

Tabela 2. Índice de Sinistralidade e exposição para os anos de 2012 a 2020, considerando veículos Passeio Nacional. Painel A - por região. Painel B - por idade do veículo.

<b>Painel A</b>		
<b>Região</b>	<b>Exposição</b>	<b>Sinistralidade</b>
MG/Metr. BH – C. Oeste – Zona Mata	4.319.823	0,72
MG/região Sul	1.015.872	0,66
RJ/Metr. Rio de Janeiro	5.120.579	0,75
RJ/Interior	1.572.079	0,69
SP/Metr. São Paulo	12.859.709	0,68
SP/Rib. Preto e demais Campinas	8.568.633	0,64
<b>Painel B</b>		
<b>Idade do Veículo (anos)</b>	<b>Exposição</b>	<b>Sinistralidade</b>
0	30.348	0,63
1	1.942.793	0,66
2	4.447.474	0,69
3	4.843.941	0,71
4	4.714.263	0,72
Maior ou igual a 5	17.477.873	0,68

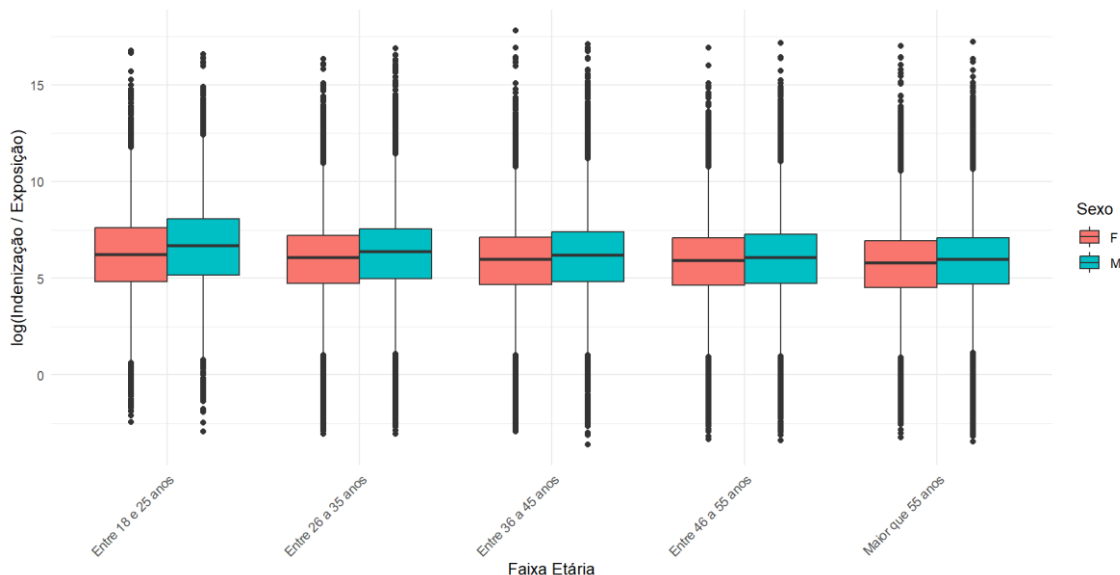
Fonte: elaborado pelos autores (2025)

As Figuras 1, 2 e 3 mostram os box-plots da variável  $\log(\text{Indenizações}/\text{Exposição})$  em relação às faixas etárias, regiões e idade do veículo, todas segregadas por sexo do condutor para os veículos do tipo passeio nacional entre os anos de 2012 e 2020. A variável exposição é uma *off-set* comum em modelos lineares generalizados (GLM) de precificação em seguros, por isso foi colocada, para fins de produção de gráfico, no denominador para relativizar os volumes de indenizações. Podemos observar que condutores do sexo masculino possuem indenizações relativas em média, e outros quantis também, maiores que do sexo feminino. Em relação à faixa etária e à região geográfica, observa-se um comportamento análogo ao da sinistralidade – quanto maior a faixa etária menor a indenização relativa à exposição e capitais possuindo essa relação maior que o interior.

Já a idade do veículo (Figura 3) aponta uma relação indenização sobre exposição diferente do índice de sinistralidade (painel B da Tabela 2). A Figura 3 mostra uma aparente queda na média da indenização relativa, enquanto o índice de sinistralidade

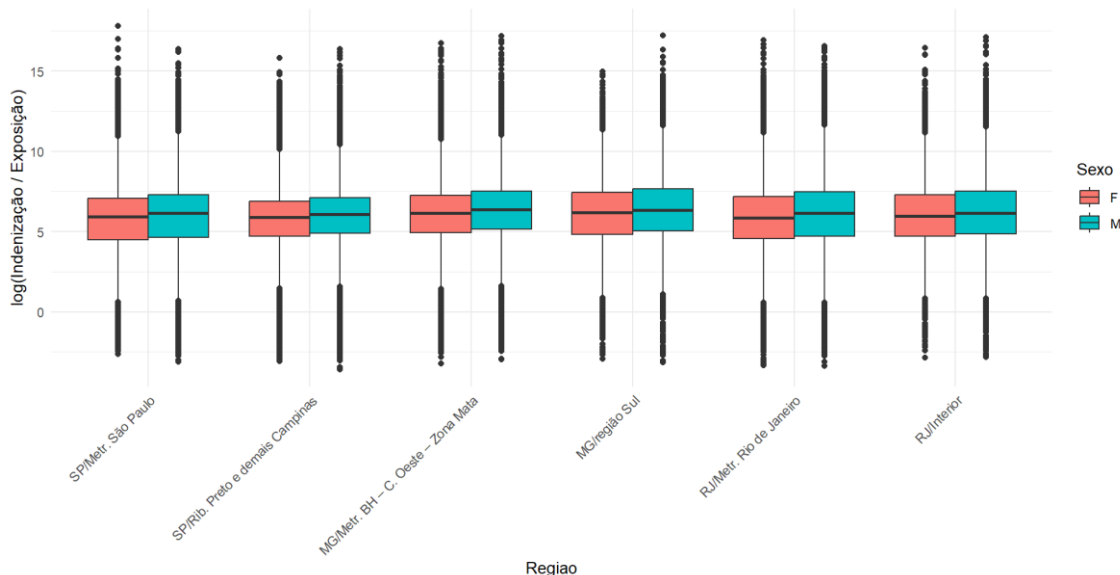
aumenta com a idade do veículo até 4 anos. Isso pode indicar que veículos mais novos possuem prêmios relativamente mais elevados, baixando o índice de sinistralidade.

Figura 1. Relação do  $\log(\text{Indenização}/\text{Exposição})$  com as faixas etárias segregado por sexo (M ou F) do condutor para os anos de 2012 a 2020, considerando veículos Passeio Nacional e nas regiões SP/Metropolitana de São Paulo; SP/Ribeirão Preto e demais municípios de Campinas; MG/região Sul; MG/Metropolitana de BH – Centro Oeste – Zona da Mata – C. Vertentes; RJ/Metropolitana do Rio de Janeiro e RJ/Interior.



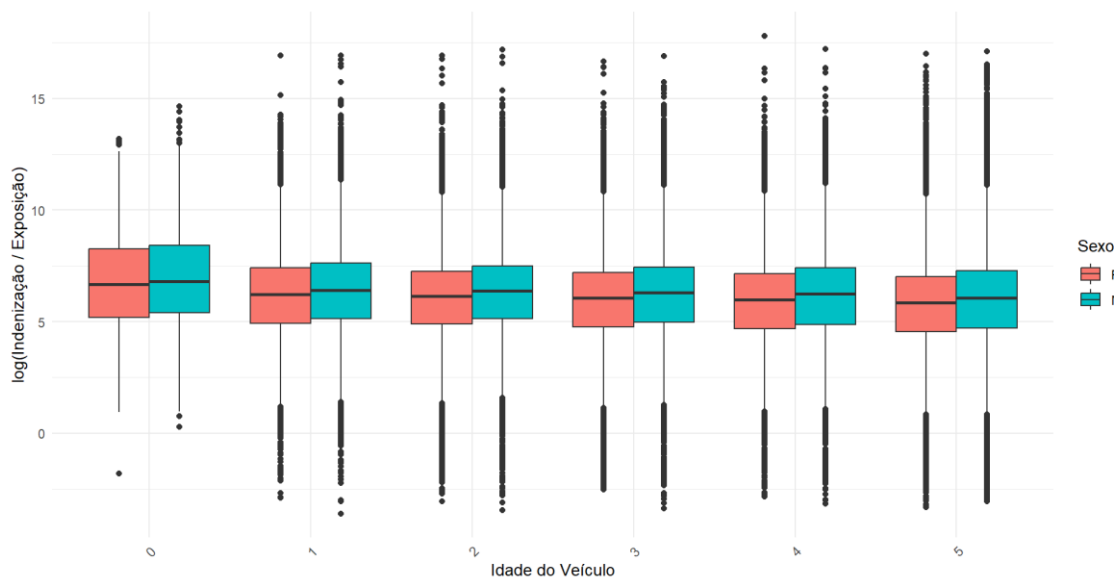
Fonte: elaborado pelos autores (2025)

Figura 2. Relação do  $\log(\text{Indenização}/\text{Exposição})$  com as regiões segregado por sexo (M ou F) do condutor para os anos de 2012 a 2020, considerando veículos Passeio Nacional.



Fonte: elaborado pelos autores (2025)

Figura 3. Relação do  $\log(\text{Indenização}/\text{Exposição})$  com a idade do veículo segregado por sexo (M ou F) do condutor para os anos de 2012 a 2020, considerando veículos Passeio Nacional e nas regiões SP/Metropolitana de São Paulo; SP/Ribeirão Preto e demais municípios de Campinas; MG/região Sul; MG/Metropolitana de BH – Centro Oeste – Zona da Mata – C. Vertentes; RJ/Metropolitana do Rio de Janeiro e RJ/Interior.



Fonte: elaborado pelos autores (2025)

## 4. Metodologia e resultados

### 4.1 Ajuste de modelos aos dados

Neste estudo, testamos (i) Modelos Lineares Generalizados – GLMs (com verossimilhança Poisson para a frequência e Gama para a severidade média) e (ii) algoritmos de machine learning *random forest*, *boosting* e *XGBoost*, tanto para a frequência quanto para a severidade média.

Utilizamos como métrica de desempenho e comparação o erro absoluto médio (MAE). Nesta subseção, comentários breves sobre cada método para previsão serão passados para que possamos focar, posteriormente, no desenvolvimento com o método mais adequado escolhido.

Historicamente, a tarifação de seguros automóvel tem se apoiado em GLM devido à sua flexibilidade e capacidade de modelar a relação entre variáveis preditivas e a resposta de interesse, geralmente número ou perdas com sinistros (MCCULLAGH e NELDER, 2019).

Neste estudo, fizemos uso do GLM com verossimilhança Poisson para a variável frequência de sinistros (FREQ) e verossimilhança Gama para a variável severidade média (resultado da divisão de INDENIZ por FREQ). Para os algoritmos de *random forest* e

*boosting*, também separamos a modelagem da frequência e da severidade média. Para o *XGBoost*, para a variável frequência foi utilizada a função de perda Poisson e para a severidade média foi utilizada a função de perda Gama.

$(Z_1, D_1, Y_1), \dots, (Z_n, D_n, Y_n)$  é uma amostra de tripletos independentes e identicamente distribuídos, em que, para a  $i$ -ésima unidade amostral,  $Z_i \in R^p$  é o vetor de preditores,  $D_i \in N$  denota a frequência de sinistros ocorridos para a  $i$ -ésima unidade amostral ( $D_i = FREQ_i$ ) e  $Y_i \in R^+$  é a severidade por sinistro da  $i$ -ésima unidade amostral ( $Y_i = \frac{IDENIZ_i}{FREQ_i}$ ). Nosso objetivo para cálculo da cota de rateio para a regra de compartilhamento da média proporcional em um arranjo mutualista é obter a perda esperada por unidade amostral (membro)  $i$ :  $E(X_i) = E(D_i)E(Y_i)$ .

Esse é o valor que desejamos prever para o ano seguinte e que determinará o rateio de cada membro de um grupo em um arranjo mutualista, conforme a regra de compartilhamento da média proporcional. O custo final que cada membro pagará é a multiplicação desse rateio pela despesa total do grupo observada no ano, conforme estabelece a equação (2) e que foi exemplificada em números no item (iv) da seção 1.2.

Como afirmamos acima, para fins de comparação entre os modelos, foi utilizada a métrica de performance considerando o MAE fora da amostra. O MAE final foi calculado como a média dos MAE anuais. O procedimento foi feito da seguinte forma: foi utilizado o primeiro ano da amostra (2012) como amostra de treino. Foram obtidos os valores preditos de 2013 e calculado o MAE em 2013 (pela diferença entre a variável *INDENIZ* e os valores preditos de  $E(X_i)$ ).

Depois foi repetido o processo: dados de 2013 como amostra de treino e calculado o MAE em 2014 e assim por diante até 2020. Ao final, o MAE do modelo é a média dos MAEs anuais.

Por mais que modelos GLM sejam simples, facilmente explicados e possuam propriedades estatísticas interessantes, eles costumam ser restritos para captarem efeitos complexos (BLIER-WONG *et al.*, 2020). Nesse sentido, o advento e a evolução das técnicas de *Machine Learning* nos últimos anos têm oferecido novas oportunidades para a precificação de seguros. A utilização de técnicas de aprendizado de máquina na precificação e análise de seguros oferece uma abordagem mais flexível e potencialmente mais precisa para identificar padrões complexos e não lineares nos dados e a sua aplicação

no contexto atuarial fez com que desenvolvesse um novo campo de estudos, os estudos em *Actuarial Learning* (TZOUGAS & KUTZKOV, 2023).

Assim, métodos como *random forest*, *boosting* e *XGBoost* têm sido explorados para melhorar a acurácia das previsões de sinistros, capturando complexidades e interações entre variáveis que métodos tradicionais podem não detectar. Segundo BLIER-WONG et al. (2020), uma vantagem significativa dos modelos recentes de *actuarial learning* (AL) é que eles aprendem transformações não lineares e interações entre variáveis dos dados sem especificá-las manualmente. Isso é realizado implicitamente com modelos baseados em árvores. De acordo com Tzougas e Kutzkov (2023), os métodos de AL que têm sido usados até agora para abordar com eficiência a regressão alternativa e problemas de classificação em seguros incluem *XGBoost*, *random forest*, *boosting*, entre outros.

Izbicki e Santos (2020) defendem que a metodologia de *random forest* tem o potencial para combinar diversas árvores de regressão que fazem a predição do comportamento de uma mesma variável explicativa. A técnica de *random forest* modifica o mecanismo de criação das árvores para que essas se tornem diferentes umas das outras. Portanto, ao invés de escolher qual das  $d$  covariáveis será utilizada em cada um dos nós da árvore, em cada passo só é permitido que seja escolhida uma dentre as  $m < d$  covariáveis. Dessa maneira, as covariáveis que entram no modelo são escolhidas de forma aleatória dentre aquelas originais e, para cada nó, um novo conjunto de covariáveis é selecionado. Em nosso estudo, para o *random forest* (tanto para a frequência quanto para a severidade média) foi usado o algoritmo de Breiman (BREIMAN, 2001). Utilizamos número de variáveis aleatoriamente amostradas como candidatos em cada divisão como sendo a raiz quadrada do número de regressores.

De forma semelhante à *random forest*, a técnica *boosting* agrega diferentes estimadores em função da regressão. No entanto, diferentemente da *random forest*, no *boosting* o estimador em função das covariáveis é construído de forma incremental. O *boosting* é um método de aprendizado de máquina que cria um modelo forte a partir de um número de modelos fracos (IZBICKI & SANTOS, 2020). No contexto da regressão, esses modelos fracos são frequentemente árvores de decisão. O modelo de *boosting* ajusta repetidamente as árvores de decisão aos resíduos (erros) da iteração anterior para melhorar a previsão. Assim como o *random forest*, não há uma fórmula algébrica simples

para a representação de um modelo de *boosting*, pois ele é construído iterativamente e pode envolver várias árvores de decisão. No entanto, de forma genérica, o modelo final de *boosting* pode ser descrito como uma soma ponderada das árvores de decisão. Em nosso estudo, utilizamos como parâmetros do *boosting* (tanto para a frequência quanto para a severidade média) árvores de regressão com profundidade máxima de 6, 1.000 interações e parâmetro de encurtamento igual a 0,50.

CHEN e GUESTRIN (2016) descreveram em seu artigo um sistema escalonável de aumento de árvore ponta a ponta chamado *XGBoost*, que é usado amplamente por cientistas de dados “para alcançar resultados de última geração em muitos desafios de aprendizado de máquina” (p. 785). Os autores demonstraram no artigo que a utilização da técnica *XGBoost* é eficiente para modelar uma ampla gama de situações. Nessa linha, a literatura afirma que o fator mais importante por trás do sucesso do *XGBoost* é sua escalabilidade em todos os cenários. O sistema funciona dez vezes mais rápido do que as soluções populares existentes em uma única máquina e pode ser dimensionada para bilhões de exemplos em sistemas distribuídos ou configurações de memória limitada. A escalabilidade do *XGBoost* se deve para vários sistemas importantes e otimizações algorítmicas (CHEN & GUESTRIN, 2016; IZBICKI & SANTOS, 2020).

Suponha que temos um conjunto de dados  $\{(z_i, d_i, y_i)\}$ , onde  $z_i$  representa as características (ou variáveis independentes) de cada observação  $i$ , e  $y_i$  (ou  $d_i$ , no caso da frequência) é a variável dependente (ou rótulo) correspondente. O objetivo do *XGBoost* é construir um modelo  $f(z)$  que preveja  $y$  (ou  $d$ ) com base em  $z$ . Para isso, ele cria uma soma de árvores de decisão em que  $k(z)$  é o número total de árvores e cada  $f_k(z)$  é uma árvore de decisão (9):

$$f(z) = \sum_{k=1}^K f_k(z) \quad (9)$$

Para a variável dependente  $y$ , a função objetivo do *XGBoost* que o algoritmo tenta minimizar é composta por uma função de perda  $L$  e um termo de regularização  $\omega$  (10):

$$Obj = \sum_i L(y_i, z_i) + \sum_k \omega(f_k) \quad (10)$$

Em que  $y_i$  é o valor observado,  $y = f(z)$  é a previsão do modelo, e  $(f_k)$  é a  $k$ -ésima árvore. O termo de regularização  $\omega$  é usado para controlar a complexidade do modelo. Durante o treinamento, o *XGBoost* atualiza os pesos das folhas de cada árvore para minimizar a função objetivo. Ele usa o gradiente da função de perda para encontrar

a direção que reduzirá o erro, ajustando-se pelos parâmetros de regularização e taxa de aprendizado.

#### 4.1.1 Modelos para a Frequência

De modo a prever a variável resposta ( $FREQ_i$ ) para o membro  $i$ , define-se o vetor de variáveis preditoras denotado por  $Z_i$ . Este vetor é composto pelas covariáveis Região, Idade do Veículo, Faixa Etária, Sexo e Exposição. Especificamente para o modelo GLM, Exposição foi utilizada como variável offset. Para o GLM Poisson, a função de ligação foi logarítmica e o modelo foi ajustado por máxima verossimilhança.

O modelo que apresentou a menor média dos MAEs anuais foi escolhido como mais adequado. Com este critério, o modelo mais adequado foi o *XGBoost*. Ele foi configurado para usar árvores de decisão com algoritmo de *booster*, especificamente com a função objetivo *Poisson*, que é adequada para variável resposta de frequência (contagem de sinistros). O modelo inclui parâmetros de regularização e controle de complexidade para otimizar o desempenho e evitar o *overfitting*. Utilizamos árvores de profundidade máxima de 8,  $\eta = 0,2$  e  $\gamma = 0,5$ . Esse conjunto de parâmetros proveu os melhores resultados de *MAE* fora-da-amostra para o *XGBoost*.  $\eta$  controla a taxa de aprendizagem – é usada para prevenir *overfitting*.  $\gamma$  é a redução da perda mínima para que seja necessária uma nova partição no nó da folha na árvore – quanto maior, mais conservador será o algoritmo.

#### 4.1.2 Modelos para a Severidade média

De modo a prever a variável resposta severidade por sinistro para o membro  $i$ , define-se o vetor de variáveis preditoras, denotado por  $Z_i$ . Este vetor é composto pelas covariáveis Região, Idade do Veículo, Faixa Etária, Sexo e Importância Segurada média. Para o modelo GLM Gama, a função de ligação foi inversa e o modelo foi ajustado por máxima verossimilhança. A variável resposta para severidade média foi construída apenas para os registros que apresentaram sinistro na amostra de treino, ou seja, quando  $FREQ_i > 0$ , e resulta da divisão entre a variável  $INDENIZ_i$  por  $FREQ_i$ .

O modelo que apresentou a menor média dos MAEs anuais foi escolhido como mais adequado. Com este critério, o modelo mais adequado foi o *XGBoost*. Ele foi configurado para usar árvores de decisão com algoritmo de *booster*, especificamente com a função objetivo *Gama*. Também utilizamos árvores de profundidade máxima de 8,  $\eta = 0,2$  e  $\gamma = 0,5$ .

O procedimento de cálculo do rateio para cada membro  $i$  em um arranjo mutualista foi:

- a. Ajuste do *XGBoost*, tanto para frequência quanto para severidade média, aos dados no ano  $t$  (onde  $t = 2012, \dots, 2019$ ) – bases de treino.
- b. Cálculo do valor predito ( $E(X_i) = E(D_i)E(Y_i)$ ) para cada registro  $i$  no ano  $t+1$  – base de teste.
- c. Utilizando a regra de compartilhamento de média proporcional, o cálculo do custeio (*ex-post*) do seguro descentralizado ( $P_i$ ) para cada registro  $i$  para o ano  $t+1$  é (11):

$$P_i = S \cdot \frac{E(X_i)}{E(S)} = S \cdot \frac{E(X_i)}{\sum_i E(X_i)} \quad (11)$$

onde  $S$  é o total de indenizações pagas no ano  $t+1$ .

#### 4.2 Comparação entre o prêmio do seguro descentralizado e o prêmio do seguro tradicional

Nesta seção, efetuamos 2 análises: (i) se há relação entre a diferença do prêmio do seguro tradicional e o do seguro descentralizado com Importância Segurada, com a Faixa Etária e o Sexo do condutor e com a Região do veículo e (ii) cálculo do carregamento máximo no prêmio puro do seguro descentralizado para igualar média dos prêmios com os do seguro tradicional.

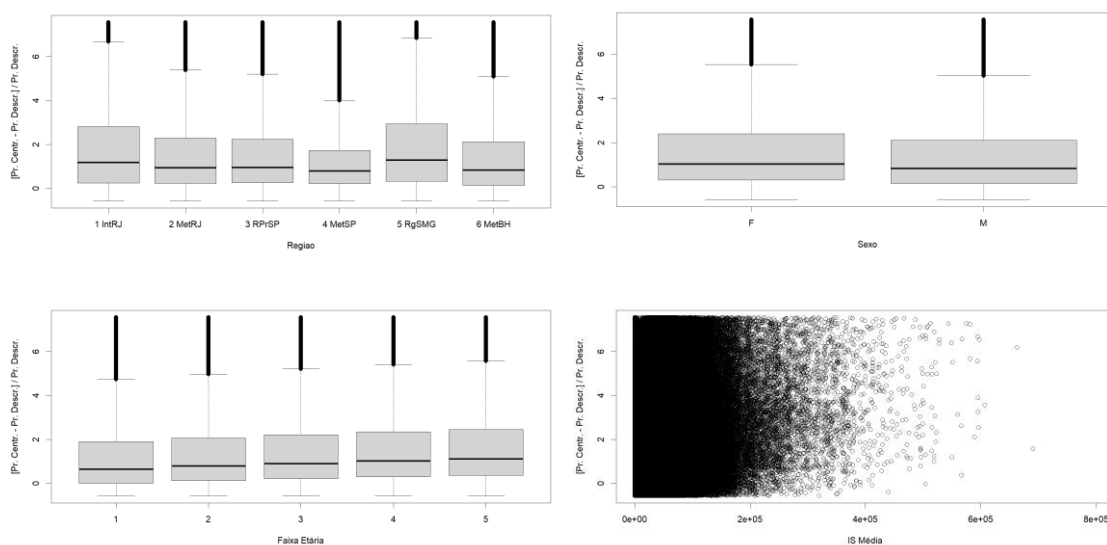
A primeira análise busca verificar se há algum perfil de apólice em que a diferença entre o prêmio do seguro tradicional para o do seguro descentralizado é maior. Nesta análise, não é relevante o valor absoluto dessa diferença. Até porque, cabe destacar, que o prêmio do seguro tradicional é um prêmio comercial e o prêmio calculado inicialmente para o seguro descentralizado é um prêmio puro. Nas seções seguintes, faremos análise do carregamento máximo que pode ser adicionado ao prêmio puro do seguro descentralizado para equivaler ao seguro tradicional, seja pela média ou pela indiferença na tomada de decisão.

Os gráficos da Figura 4 mostram a relação entre a diferença entre o prêmio do seguro tradicional e o do seguro descentralizado com a Importância Segurada, Sexo, Faixa Etária e Região.

Os gráficos da Figura 4 da diferença entre os prêmios em função da Importância Segurada e do sexo do condutor parecem não indicar uma relação entre essas variáveis. Já os gráficos da faixa etária do condutor e a região indicam alguma relação. Podemos

observar que a faixa etária dos motoristas mais jovens possui uma concentração de diferenças maiores entre os prêmios do que as demais faixas. Adicionalmente, para as regiões de interior dos estados do RJ, SP e MG também observamos uma concentração de diferenças maiores dos prêmios do que nas capitais. Não podemos esquecer que a comparação neste momento é entre um prêmio comercial efetivamente cobrado no seguro tradicional e o que seria o prêmio puro de um seguro descentralizado. Portanto, essas diferenças englobam também políticas distintas para carregamento do prêmio puro por parte das seguradoras entre 2012 e 2020. Por exemplo, é possível que as regiões de interior de RJ, MG e SP possuem carregamentos (de segurança, custo de aquisição, despesas administrativas ou margem de lucro) maiores que nas capitais, talvez por questões de concorrência maior nestas últimas regiões. Como não é possível saber a composição do preço final do seguro cobrado ao segurado, não se sabe o carregamento implícito no preço, ou seja, o seu valor e nem se é positivo ou até mesmo negativo.

Figura 4. Gráficos da diferença entre o prêmio comercial do seguro tradicional e o prêmio puro do seguro descentralizado pela (i) região; (ii) sexo do condutor; (iii) faixa etária do condutor e (iv) Importância Segurada.



Fonte: elaborado pelos autores (2025)

Portanto, há duas possibilidades para estas observações: (i) ou a diferença entre o prêmio do seguro tradicional e o do descentralizado tende de fato a ser maior para motoristas mais jovens e no interior destes estados para a amostra de apólices entre os anos de 2012 e 2020 ou (ii) o nicho de motoristas mais jovens e as regiões de interior

destes estados possuem um carregamento dentro do preço do seguro tradicional maior. E esse carregamento pode ser maior por motivos de concorrência, ou por qualquer uma das teóricas parcelas do carregamento (de segurança ou custo de aquisição/comissão ou DA ou margem de lucro).

#### 4.2.2 Cálculo do carregamento máximo no seguro descentralizado para igualar médias dos prêmios com os do seguro tradicional

Em posse dos prêmios puros calculados para o seguro descentralizado e com os dados de prêmio comercial do seguro tradicional presentes no banco de dados, calculamos qual seria o carregamento a ser recolhido sobre o prêmio puro do seguro descentralizado de forma a igualar, na média, ambos os tipos de seguro. Podemos efetuar esse cálculo por região, por ano, por sexo, por faixa etária, enfim por qualquer característica das apólices presentes no banco de dados. Para todos os registros (todos os anos), o carregamento sobre o prêmio puro do seguro descentralizado que iguala, na média, com o prêmio do seguro tradicional é igual a 39,3%.

Tabela 3. Carregamento sobre o prêmio puro do seguro descentralizado que iguala, na média, com o prêmio do seguro tradicional de veículos da categoria passeio nacional, por região, para os anos de 2012 a 2020.

Região	Carregamento (%)
MG/Metr. BH – C. Oeste – Zona Mata	28,9
MG/região Sul	46,7
RJ/Metr. Rio de Janeiro	27,3
RJ/Interior	32,9
SP/Metr. São Paulo	30,6
SP/Rib. Preto e demais Campinas	37,0

Fonte: elaborado pelos autores (2025)

## 5. Considerações finais

O setor de seguros vem passando por transformações marcantes nos últimos tempos. Uma destas é a introdução da chamada proteção patrimonial mutualista, regulamentada pela recém-publicada Lei Complementar n. 213/2025. Em comparação com o seguro “tradicional”, a proteção mutualista possui diferenças, em especial, no que se refere ao custeio. Esse é baseado em um mecanismo de compartilhamento de risco, onde os participantes de um grupo compartilham os riscos entre eles pagando *ex-post* uma contribuição para o grupo que, agregada, é responsável por cobrir toda a perda observada dos membros. Essas contribuições seguem uma regra de compartilhamento de risco

estabelecida previamente. A regra usada neste artigo e a mais observada na prática é a da média proporcional.

Há um pensamento em atores do mercado de seguros de que este tipo de cobertura é de menor preço que o seguro tradicional. Neste artigo, procuramos endereçar essa questão para o seguro automóvel, que é uma linha de negócio de muita relevância, além de ser o principal mercado para as associações que atuam com o modelo de proteção patrimonial mutualista. Dessa forma, usando dados reais e públicos de apólices subscritas por seguradoras entre os anos de 2012 e 2020 (último ano disponível dos dados), comparamos os custos do seguro automóvel na abordagem descentralizada com o prêmio efetivo do seguro tradicional (informação presente no banco de dados). Essa comparação foi feita por meio do cálculo do carregamento que adicionado ao prêmio puro da proteção mutualista o deixa em equilíbrio, na média, com o do seguro tradicional.

De forma geral, nossos resultados indicam que o carregamento sobre o prêmio do seguro descentralizado (proteção mutualista) pode ser da ordem de 25% a 40% que ainda assim seria menor que o prêmio do seguro tradicional. Tais valores dependem da característica do risco, do nível de concorrência em cada segmento/nicho, entre outros. Para nichos/segmentos de mercado com menores margens nos prêmios do seguro tradicional, ou seja, em teoria segmentos com maior concorrência, há menos espaço para margens sobre o custo da proteção patrimonial mutualista.

A presente pesquisa vem em momento crucial para o mercado e busca endereçar a comparação no custeio destes dois tipos de operação de seguro.

## Referências

- AKERLOF, G. A. The market for “lemons”: Quality uncertainty and the market mechanism. **Decision Science**, v. 84, n. 3, p. 261-273, 2017.
- ARNOTT, R. J.; STIGLITZ, J. E. **The basic analytics of moral hazard**. 1988.
- BRASIL**. Lei Complementar nº 213, de 15 de janeiro de 2005. Dispõe sobre o Sistema Nacional de Seguros Privados. **Diário Oficial da União**, Brasília, DF, janeiro de 2005.
- BLIER-WONG, C. et al. Machine Learning in P&C Insurance: A Review for Pricing and Reserving. **Risks**, v. 9, n. 1, p. 4, 23 dez. 2020.
- BORSELLI, A. Smart contracts in insurance: a law and futurology perspective. In: **InsurTech: A Legal and Regulatory View**. Cham: Springer International Publishing, 2019. p. 101-125.
- BREIMAN, L. Random Forests. **Machine Learning**, vol. 45, no. 1, pp. 5–32, 2001.

BUTERIN, V. A next-generation smart contract and decentralized application platform. **Ethereum White Paper**, 2014. Disponível em: <https://ethereum.org/en/whitepaper/>. Acesso em: 10 out. 2023.

CNSP, 2021. Resolução CNSP n. 432, de 12 de novembro de 2021. Disponível em: [www.susep.gov.br](http://www.susep.gov.br).

CHEN, T.; GUESTRIN, C. *XGBoost*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. **Anais**. New York, NY, USA: ACM, 13 ago. 2016.

DHAENE, J. Decentralized risk-sharing: definitions, properties and characterizations – Pooling risks without a central insurer. Short course in IME congress at Chicago, Illinois, July 8, 2024.

FENG, R. **Decentralized insurance: Technical Foundations of Business Models**. Springer. 2023.

FERREIRA, P.P. **Modelos de Precificação e Ruína de Seguros de Curto Prazo**. Escola Nacional de Seguros. 2002.

FERREIRA, P. P; MANO, C. M. C. A. **Aspectos Atuariais e Contábeis das Provisões Técnicas**. Escola Nacional de Seguros. 2009.

GRACE, M. F. Regulation of Insurance Markets in the USA. **In: Handbook of Insurance: Volume II**. Cham: Springer Nature Switzerland, 2025. p. 457-482.

HENCKAERTS, R.; CÔTÉ, M.-P.; ANTONIO, K.; VERBELEN, R. Boosting Insights in Insurance Tariff Plans with Tree-Based Machine Learning Methods. **North American Actuarial Journal**, vol. 25, no. 2, pp. 255–285, 2021

IZBICKI, R.; SANTOS, T. M. **Aprendizado de máquina: uma abordagem estatística**. 1ª edição ed. São Paulo- SP: Câmara Brasileira do Livro, 2020.

KAAS, R.; GOOVAERTS, M.; DHAENE, J. **Modern actuarial risk theory**. Springer Science & Business Media, 2008.

KITCHENS, F. L. Financial implications of artificial Neural Networks in automobile insurance underwriting. **International Journal of Electronic Finance**, v. 3, n. 3, p. 311-319, 2009.

KUNREUTHER, H. C.; PAULY, Mark V.; MCMORROW, Stacey. *Insurance and behavioral economics: Improving decisions in the most misunderstood industry*. Cambridge University Press, 2013.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. [s.l.] Routledge, 2019.

SCHÄR, F. Decentralized finance: on blockchain and smart contract-based financial markets. **Review of the Federal Reserve Bank of St Louis**, v. 103, n. 2, p. 153-174, 2021.

SUSEP, 2021. Circular Susep n. 639, de 9 de agosto de 2021. Disponível em: [www.susep.gov.br](http://www.susep.gov.br).

SZABO, N. Formalizing and securing relationships on public networks. **First monday**, 1997.

RUBINSTEIN CAVALCANTI, L. **A business model for vehicle insurance based on blockchain smart-contracts**. 2018. Tese de Doutorado. Politecnico di Torino.

SU, X.; BAI, M. Stochastic gradient boosting frequency-severity model of insurance claims. **PLoS one**, vol. 15, no. 8, p. e0238000, 2020.

TZOGAS, G.; KUTZKOV, K. Enhancing Logistic Regression Using Neural Networks for Classification in Actuarial Learning. **Algorithms**, v. 16, n. 2, p. 99, 9 fev. 2023.

WERBACH, K. **The blockchain and the new architecture of trust**. MIT Press, 2018.

ZETZSCHE, D. A.; ARNER, D. W.; BUCKLEY, R. P. Decentralized finance. **Journal of Financial Regulation**, v. 6, n. 2, p. 172-203, 2020.

## MUTUAL AID: A COST COMPARISON WITH TRADITIONAL INSURANCE

### Abstract

*The insurance sector in Brazil has been undergoing profound transformations in recent years. Among these developments is the introduction of mutual aid protection, regulated by a recently enacted law. In comparison with “traditional” insurance, mutual aid exhibits notable differences, particularly with respect to its funding structure. In international markets, this arrangement is often referred to as decentralized insurance. Unlike traditional insurance, in which premiums are determined ex-ante, mutual protection establishes the “premium” ex-post, through the apportionment of claim-related expenses across the pool of participants. Employing a machine learning model — identified as the most effective in terms of predictive performance — this study analyzes automobile insurance costs using real and publicly available data from policies underwritten by insurers between 2012 and 2020, the most recent year for which data is available. Furthermore, we estimate the maximum premium loading that administrators of mutual aid could have imposed on these policies while remaining, on average, as attractive to policyholders as traditional insurance. Finally, by applying an exponential utility framework in conjunction with conformal inference techniques, we compute the level of premium loading under mutual asset protection that renders policyholders indifferent between this mechanism and traditional insurance. This research is presented at a pivotal juncture for the insurance industry and aims to contribute to the ongoing discussion on the funding mechanisms underlying these two distinct forms of risk transfer.*

**Keywords:** *Mutual Aid; Decentralized Insurance; Motor Insurance; Pricing.*