UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE CIÊNCIAS CONTÁBEIS E ATUARIAIS
CURSO DE CIÊNCIAS ATUARIAIS

THAMIRIS DOMINGOS DA SILVA

ANÁLISE TÉCNICA DA CARTEIRA DE SEGUROS HABITACIONAIS E DE RESPONSABILIDADE CIVIL: Sinistralidade, Segmentação por Risco e Equidade na Precificação

Recife
2025

THAMIRIS DOMINGOS DA SILVA

ANÁLISE TÉCNICA DA CARTEIRA DE SEGUROS HABITACIONAIS E DE RESPONSABILIDADE CIVIL: Sinistralidade, Segmentação por Risco e Equidade na Precificação

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Ciências Atuariais do Campus Recife da Universidade Federal de Pernambuco, na modalidade de monografia, como requisito parcial para obtenção do grau de bacharel em Ciências Atuariais.

Orientador: FILIPE COSTA DE SOUZA

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Silva, Thamiris Domingos da .
Análise técnica da carteira de seguros Habitacionais e de Responsabilidade Civil: sinistralidade, segmentação por risco e equidade na precificação / Thamiris Domingos da Silva. - Recife, 2025.
74, tab.

Orientador(a): Filipe Costa de Souza
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Ciências Sociais Aplicadas, Ciências Atuariais, 2025.
Inclui referências, apêndices.

1. Seguro Habitacional. 2. Seguro de Responsabilidade Civil. 3. Sinistralidade . 4. Precificação. 5. Seleção adversa. 6. Risco moral. I. Souza, Filipe Costa de. (Orientação). II. Título.

310 CDD (22.ed.)

THAMIRIS DOMINGOS DA SILVA

ANÁLISE TÉCNICA DA CARTEIRA DE SEGUROS HABITACIONAIS E DE RESPONSABILIDADE CIVIL: Sinistralidade, Segmentação por Risco e Equidade na Precificação

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Ciências Atuariais do Campus Recife da Universidade Federal de Pernambuco, na modalidade de monografia, como requisito parcial para obtenção do grau de bacharel em Ciências Atuariais.

Aprovada em: 03/12/2025

BANCA EXAMINADORA

gov.br Documento assinado digitalmente
FILIPE COSTA DE SOUZA
Data: 04/12/2025 13:42:19-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Filipe Costa de Souza (Orientador)
Universidade Federal de Pernambuco

gov.br Documento assinado digitalmente
RENATA GOMES ALCOFORADO
Data: 03/12/2025 17:18:14-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. Renata Gomes Alcoforado
Universidade Federal de Pernambuco

Assinado por: Alfredo Duarte Egídio dos Reis
Identificação: BI04573906
Data: 2025-12-03 às 20:39:36

Prof. Dr. Alfredo Duarte Egídio dos Reis
ISEG – Universidade de Lisboa

Dedico este trabalho aos meus pais, que sonharam comigo, acreditaram e nunca me deixaram desistir.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por toda força, coragem e resiliência nesta jornada.

Aos meus pais, que sempre fizeram o possível para tornar essa fase mais leve, me amaram e acolheram nos momentos mais difíceis e foram minha força diária para chegar até aqui.

À minha irmã, que sempre esteve ao meu lado nos ruins e bons momentos.

À minha família, pelo apoio, paciência e amor.

À Professora Renata Alcoforado, que com toda a sua bondade e empatia me acolheu e orientou. Obrigada por além de compartilhar seu conhecimento, sempre demonstrar compreensão, apoio e motivação. Deu certo.

Aos meus amigos, por cada conversa, risada, desabafo e incentivo. Vocês deram cor aos meus dias cinzentos.

Aos meus amigos do trabalho, que tornam meus dias mais leves apesar da pressão diária, por compartilhar conhecimentos e me incentivarem. Em especial, ao meu gestor Joabe, pela paciência e compreensão nos momentos em que precisei me dedicar mais a este trabalho.

Aos amigos da graduação, pelo companheirismo, pelas trocas e apoio mútuo.

Por fim, agradeço aos professores e à Universidade Federal de Pernambuco pela oportunidade de acesso à educação pública e de qualidade.

“Long story short, I survived”

- Taylor Swift

RESUMO

Este trabalho analisa uma carteira de seguros Habitacionais e de Responsabilidade Civil no período 2015–2025, com objetivo de responder a seguinte questão: “existe coerência entre os prêmios arrecadados e os riscos efetivamente assumidos na carteira de uma seguradora?”. As bases de prêmios e sinistros foram integradas pela chave apólice-endosso, gerando indicadores de sinistralidade, frequência e severidade. Aplicaram-se análises descritivas, análise de cluster, testes não paramétricos, correlação de Spearman por estratos homogêneos (ramo × Unidade Federativa) e regressão logística para a ocorrência de sinistro. Os resultados mostram forte assimetria e concentração: o ramo de Responsabilidade Civil Profissional concentra a maior parte das apólices, mas baixa participação em sinistros, enquanto o Habitacional Prestamista apresenta o movimento inverso. A segmentação identificou três perfis técnico-atuariais: um grupo majoritário de baixo valor e baixa frequência, um grupo intermediário e um grupo de grande porte com alta exposição e materialidade econômica. Não se observaram padrões consistentes de subscrição adversa ou risco moral. Os dois clusters minoritários concentram contratos de maior porte, compatíveis com a natureza coletiva de parte do Habitacional. A análise de equidade apontou dispersões de prêmio muito elevadas dentro de alguns estratos (medidas de dispersões acima do esperado) e, embora a correlação prêmio-sinistro seja majoritariamente positiva, há casos pontuais de correlação fraca ou negativa que demandam revisão tarifária. Recomenda-se monitoramento por clusters, padronização tarifária nos estratos com maior dispersão e incorporação de métricas adicionais de exposição para refinar a precificação.

Palavras–chave: seguro habitacional; responsabilidade civil; precificação; sinistralidade; seleção adversa; risco moral.

ABSTRACT

This study analyzes a portfolio of Housing Insurance and Liability Insurance over the period 2015–2025, with the objective of addressing the following question: “Is there coherence between the premiums collected and the risks effectively assumed in an insurer’s portfolio?” Premium and claims databases were integrated using the policy–endorsement key, allowing for the construction of indicators of loss ratio, frequency, and severity. The analysis applies descriptive statistics, cluster analysis, nonparametric tests, Spearman rank correlation by homogeneous strata (line of business × state), and logistic regression for claim occurrence. The results reveal strong asymmetry and concentration. The Professional Liability line concentrates most policies but accounts for a low share of claims, whereas Housing Credit Insurance exhibits the opposite pattern. The segmentation identified three technical–actuarial profiles: a dominant group characterized by low premium values and low claim frequency, an intermediate group, and a large-scale group with high exposure and significant economic materiality. No consistent evidence of adverse selection or moral hazard was observed. The two minority clusters concentrate higher-value contracts, consistent with the collective nature of part of the housing insurance portfolio. The equity analysis indicates very high premium dispersion within certain strata (dispersion measures above expected levels), and although the premium–claim correlation is predominantly positive, there are isolated cases of weak or negative correlation that call for tariff revision. The study recommends cluster-based monitoring, tariff standardization in strata with greater dispersion, and the incorporation of additional exposure metrics to refine pricing.

Keywords: housing insurance; liability insurance; pricing; loss ratio; adverse selection; moral hazard.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

GLOSSÁRIO

AUC – Área Sob a Curva

PREMIT – Relatório de Prêmios Emitidos no Mês de Competência

ROC – Característica de Operação do Receptor

SINPAG – Sinistros Pagos

SINPEND – Sinistros Pendentes

UF – Unidade Federativa

SUMÁRIO

1. INTRODUÇÃO

A história do seguro está relacionada ao desenvolvimento das civilizações e à busca por mecanismos de proteção contra perdas e incertezas. O Código de Hamurabi (c. 1750 a.C.) foi decretado na Babilônia Antiga estabelecendo normas que determinavam a divisão dos prejuízos entre comerciantes, em casos de roubo ou perdas durante o transporte fluvial de mercadorias. Em seguida, na Grécia Antiga, a Lei de Rodes estabeleceu o princípio da avaria grossa, no qual os danos sofridos voluntariamente para salvar uma embarcação seriam compartilhados por todos os envolvidos na carga, antecipando os conceitos fundamentais do seguro marítimo (RIBEIRO, 1994).

Outro fato importante no surgimento do conceito de seguros ocorreu nas rotas comerciais do Oriente Médio e Norte da África, quando caravanas e camelos cruzavam desertos transportando mercadorias. Os comerciantes rateavam os custos de reposição entre si, como uma forma inicial de mutualismo e proteção coletiva, para reduzir os riscos de saques, perdas ou mortes de animais durante essas viagens. Essas práticas evoluíram gradativamente e foram formalizadas, principalmente nas cidades portuárias da Europa medieval, como Gênova e Veneza, onde surgiram os primeiros contratos de seguro por escrito. Dessa forma, o seguro se consolidou historicamente como um instrumento essencial de estabilização econômica, solidariedade e gestão de riscos (RIBEIRO, 1994).

No Brasil, as companhias de seguros surgiram com a chegada da corte portuguesa ao Rio de Janeiro em 1808, atuando no ramo de seguros marítimos. Durante a Primeira República (1889–1914), as companhias de seguro brasileiras passaram a atuar também como agentes financeiros estratégicos, comprando títulos da dívida pública e financiando o Estado. Nesse período, o setor passou a se incluir de forma ativa ao sistema financeiro e à reprodução do capital, consolidando-se como elemento importante da economia brasileira (MAGALHÃES, 1997; LANNA; SAES, 2020).

Segundo o relatório Global Insurance Market Trends 2024 da Organisation for Economic Co-operation and Development (OCDE), que analisa o desempenho do mercado segurador global com base em dados de 2023, a área apresentou crescimento sólido, porém permanece concentrado em economias avançadas, como EUA, Reino Unido, França e Japão, onde a penetração (proporção de prêmios emitidos sobre o PIB) ultrapassa 10% do PIB. Os

prêmios para os seguros não vida tiveram um aumento significativo. Os prêmios para os ramos não vida cresceram em média 12,4% em termos nominais e 6,2% em termos reais. Porém, esse crescimento se motiva principalmente pelo aumento da inflação, afetando os custos de sinistros, operacionais e taxas de resseguros.

A Superintendência de Seguros Privados (SUSEP) afirma que no Brasil o mercado de seguros manteve um forte crescimento, arrecadando o valor de R$ 388,03 bilhões, um aumento de 9% em relação a 2022. Deste total, R$125,88 bilhões são referentes aos ramos não vida e R$215,02 bilhões ao ramo vida. Os segmentos que tiveram maior índices de vendas foram o de vida e previdência, refletindo maior procura por proteção e investimento por parte da sociedade brasileira (SUSEP, 2025). Diante da magnitude desse mercado, torna-se imperativo investigar os seus mecanismos de funcionamento, sobretudo os relacionados à eficiência técnica, à precificação e à gestão de riscos.

Neste contexto, Shi e Zhao (2019) analisaram dados de seguros patrimoniais utilizando modelos de regressão de cópula, identificando dependência negativa significativa entre frequência e severidade, ou seja, um aumento no número de sinistros tende a reduzir o valor médio por sinistro. De forma complementar, Alcoforado *et al*. (2025), ao analisarem dados reais de seguros habitacionais e de responsabilidade civil, identificaram dependência estatisticamente significativa entre frequência e severidade dos sinistros, o que impacta diretamente a modelagem atuarial e a gestão de riscos.

Esses achados reforçam a necessidade de investigar se a relação entre os prêmios pagos pelos segurados e os sinistros registrados revelam possíveis ineficiências técnicas ou injustiças na precificação. Em determinadas situações, apólices com prêmios elevados não geram sinistros, enquanto outras, com prêmio mais baixo, resultam em altos valores indenizados. Essa oscilação pode ser indicativa de falhas na subscrição, risco moral, ou nos critérios de precificação adotados pela seguradora. Nesse contexto, surge a questão central que orienta este estudo: existe coerência entre os prêmios arrecadados e os riscos efetivamente assumidos na carteira de uma seguradora?

Para realizar esta investigação, o presente trabalho tem como objetivo geral analisar a relação entre os prêmios recebidos e os sinistros pagos em uma carteira de seguros nos ramos de Responsabilidade Civil e Habitacional, com foco na eficiência técnica, nos critérios de subscrição e na justiça na precificação. Para isso, utilizamos uma base de dados que consiste nos prêmios emitidos e sinistros ocorridos no período de janeiro de 2015 a julho de 2025.

Portanto, propõe-se: (i) Descrever a carteira no período 2015–2025, caracterizando prêmios, sinistros, sinistralidade, frequência e severidade por ramo e UF, e identificar padrões de assimetria e concentração [Ver 4.1]; (ii) segmentar a carteira em perfis técnico-atuariais por meio de clusterização, comparar os grupos quanto a volume, frequência, severidade e sinistralidade e verificar sinais de subscrição adversa ou risco moral [Ver 4.2]; (iii) avaliar a equidade de precificação em estratos homogêneos por ramo e UF, mensurando dispersão de preços e a coerência preço–risco via correlação de Spearman entre prêmio e sinistro (apesar o prêmio não ser uma variável aleatória, foi utilizado como *proxy* da importância segurada) [Ver 4.3]; e (iv) estimar a probabilidade de ocorrência de sinistro por regressão logística com log do prêmio e fatores de ramo e UF, avaliando desempenho discriminatório e diferenças estruturais [Ver 4.4].

Este estudo tem relevância tanto acadêmica quanto aplicada. Do ponto de vista teórico, contribui para a literatura que discute a dependência entre frequência e severidade de sinistros, a eficiência técnica das apólices e a justiça na precificação. A evidência de dependência estatística entre frequência (número de sinistro por apólice) e severidade (valor do sinistro), como constatado por Alcoforado *et al.* (2025), aponta para a necessidade de modelos mais realistas e adaptados à prática de mercado. Do ponto de vista prático, esta pesquisa oferece às seguradoras uma análise que pode motivar decisões de gestão de risco, subscrição e precificação. Além disso, a análise também pode servir como base para a proposição de melhorias nos processos de aceitação de propostas e no desenvolvimento de políticas de fidelização e segmentação de clientes.

O restante do trabalho está organizado como segue: no Capítulo 2 é visto o referencial teórico, no Capítulo 3 apresentamos a metodologia, no Capítulo 4 estão as análises e resultados, e, por fim, no Capítulo 5 encerramos com as conclusões.

2. REFERENCIAL TEÓRICO

Neste capítulo são apresentados os principais conceitos e abordagens teóricas relacionados ao contexto dos seguros. A seção explora sua subscrição de risco, precificação, sinistralidade e sua interpretação técnica, subscrição adversa, risco moral e equidade. Por fim, apresenta-se uma síntese de estudos que utilizam métodos estatísticos aplicados à análise de dados de seguros, estabelecendo conexões entre a teoria e as estratégias empíricas adotadas neste trabalho. Essa fundamentação teórica servirá de apoio para as análises desenvolvidas nos capítulos empíricos do trabalho.

2.1. Subscrição de risco

A subscrição se dá pelo processo de análise do risco, em que se avalia e decide a aceitação, estabelecendo os critérios de precificação e as condições contratuais. Rejda e McNamara (2017) explicam que a atividade deve seguir uma política de subscrição definida previamente pela empresa de acordo com os seus objetivos, os quais se baseiam em três princípios centrais, sendo eles: 1) obter lucro, produzindo uma carteira de negócios saudável e lucrativa; 2) Refinar a seleção de segurados conforme os padrões de subscrição da empresa, evitando a seleção adversa, por exemplo; e 3) garantir equidade entre os segurados, através da cobrança de taxas equitativas de acordo com seu grupo de classificação.

Para Randall (2000, *apud* Lima, 2008, p. 31), subscrever o risco permite determinar o objeto, as condições, e o preço que possibilite a seguradora manter uma carteira de negócios crescente e lucrativa. Além disso, os resultados da subscrição podem ser representados por índices, podendo ser o índice de sinistralidade, índice de despesas ou índice combinado, o qual mede a eficiência da seguradora a atividade.

Ao estudar a qualidade de subscrição de riscos das seguradoras brasileiras, Lima (2008) observou que seguradoras que possuem subscrição de riscos eficientes, alcançam melhores avaliações qualitativas no mercado. Porém, o Brasil ainda não possui uma cultura de subscrição consolidada, destacando-se a falta de profissionais atuariais na área, o que impacta na precificação das operações das seguradoras, em que seus subscritores tendem a reduzir seus prêmios para acompanhar a concorrência no mercado.

2.2. Precificação

A precificação de seguros é o processo de determinação do prêmio que o segurado deve pagar para transferir o risco à seguradora. Esse valor deve ser suficiente para cobrir o custo esperado dos sinistros, as despesas operacionais, a margem de lucro e as provisões técnicas necessárias.

Na tarifação de seguros de curto prazo, o cálculo do prêmio inicia-se pela estimativa do valor esperado das indenizações totais ocorridas, denominado prêmio de risco, que representa o custo médio do risco assumido. A este valor acrescentam-se carregamentos de segurança, o qual funciona como uma reserva adicional destinada a absorver as variações estatísticas do risco, reduzindo a probabilidade de que o valor dos sinistros ultrapasse o prêmio puro. Assim, o prêmio final, chamado prêmio comercial, resulta da soma do custo esperado do risco com os acréscimos necessários para assegurar a solvência e a estabilidade financeira da seguradora (FERREIRA, 2002).

Conforme destaca Rocha (2015), o uso de Modelos Lineares Generalizados (GLMs) tem se consolidado como uma das metodologias mais eficazes nesse processo, sobretudo por sua flexibilidade em lidar com distribuições que se afastam das premissas do modelo linear clássico, como a normalidade e a homoscedasticidade. Ainda segundo o autor, os GLMs possibilitam modelar adequadamente tanto a frequência quanto a severidade dos sinistros, utilizando distribuições como a Poisson e a Gaussiana Inversa, o que permite estimar o prêmio de risco com maior precisão. Essa abordagem contribui diretamente para a diferenciação tarifária entre segurados, proporcionando a cobrança de prêmios proporcionais ao risco individual, o que aumenta a competitividade da seguradora no mercado e reduz o risco de seleção adversa (ROCHA, 2015).

De acordo com Zhang e Walton (2019), métodos adaptativos como os Modelos Lineares Generalizados (GLMs) e a Regressão com Processo Gaussiano são mais eficientes do que as abordagens tradicionais de precificação estática, sobretudo em ambientes marcados por incerteza e competição. Os autores destacam que esses modelos, ao aprenderem continuamente com os dados observados, conseguem melhorar suas estimativas de demanda e risco ao longo do tempo, reduzindo o chamado “regret cumulativo”, ou seja, a perda causada por não se adotar o preço ótimo desde o início.

2.3. Sinistralidade e sua interpretação técnica.

A razão sinistro/prêmio, também conhecida como sinistralidade ou loss *ratio*, compõe um dos principais indicadores de desempenho técnico e financeiro das seguradoras. De forma geral, representa a proporção entre o total de sinistros incorridos e o volume de prêmios efetivamente ganhos no período. De acordo com Wongsuwatt *et al*. (2021), esse índice demonstra o percentual da receita de prêmios destinado ao pagamento de sinistros, sendo amplamente utilizado para avaliar a eficiência das práticas de subscrição e precificação de riscos. De maneira formal, a sinistralidade é calculada pela seguinte expressão (Equação 1):

$$Sinistralidade = \frac{Sinistros\ Incorridos\ Líquidos}{Prêmios\ Ganhos\ Líquidos}$$

Essa definição é atestada pela *American Academy of Actuaries* (1998), que descreve a razão sinistro/prêmio como a relação entre os sinistros ocorridos e os prêmios ganhos, ajustada, quando necessário, pelas variações nas reservas técnicas. O relatório enfatiza que sua interpretação deve ser feita com cautela, uma vez que o resultado é sensível a diversos fatores, como a estrutura de despesas, as práticas de subscrição, o método de contabilização de reservas e a composição do portfólio de negócios.

De maneira complementar, o Corporate Finance Institute (2024) apresenta uma definição incluindo as despesas de regulação e liquidação de sinistros (Equação 2), de modo que:

$$Sinistralidade = \frac{Sinistros\ Pagos + Despesa\ de\ Regulação\ de\ Sinistros}{Prêmios\ Ganhos} \times 100$$

A inclusão dessas despesas permite uma visão mais abrangente da eficiência operacional da seguradora, especialmente em ramos de seguros que envolvem elevados custos administrativos e jurídicos no processo de liquidação. Ao observar a faixa de valores obtidos pelo cálculo da taxa de sinistralidade, pode-se interpretar que os índices de sinistralidade menores do que 100% demonstram que a empresa, mesmo após o pagamento do sinistro, retém parte do prêmio adquirido; sinistralidade iguais a 100% mostram que a empresa não está lucrando e nem tendo prejuízo e, por fim, quando maiores que 100%, expressa prejuízo, ao passo que quanto maior o índice de sinistralidade, maior sua perda.

No contexto dos seguros não vida, Wongsuwatt et al. (2021) analisaram empiricamente a influência da sinistralidade sobre a lucratividade de 52 seguradoras tailandesas entre 2016 e 2018, concluindo que há uma relação negativa e estatisticamente significativa entre o nível de

sinistralidade e indicadores de rentabilidade, como o retorno sobre ativos (ROA), o retorno sobre o patrimônio líquido (ROE) e as margens de lucro. Segundo os autores, valores persistentemente elevados da sinistralidade indicam deterioração da rentabilidade e podem refletir falhas de precificação, aumento da frequência de sinistros ou ineficiências operacionais. Por outro lado, índices moderados expressam equilíbrio técnico, pois demonstram que a seguradora é capaz de cobrir seus compromissos com sinistros, despesas administrativas e ainda manter uma margem de lucro sustentável.

O relatório da *American Academy of Actuaries* (1998) complementa essa análise ao demonstrar que, apesar de amplamente adotado por reguladores e companhias, o *loss ratio* não deve ser utilizado isoladamente como medida de rentabilidade. O documento aponta que políticas regulatórias que fixam padrões mínimos de sinistralidade podem gerar distorções, pois desconsideram as diferenças estruturais entre produtos, despesas e carteiras de risco. Assim, recomenda-se que a avaliação da eficiência técnica seja acompanhada por indicadores de solvência, retorno sobre o capital e aderência às hipóteses atuariais.

Grace (2021), ao estudar a dinâmica da *loss ratio* no mercado de seguros de propriedade e responsabilidade civil nos Estados Unidos, identificou uma forte associação entre o comportamento desse indicador e as condições macroeconômicas e à capacidade de capital das seguradoras. Utilizando modelos de Markov com mudança de regime, o autor identificou dois estados predominantes, um de baixa sinistralidade e alta rentabilidade, e outro de alta sinistralidade e margens reduzidas, porém não foi encontrada evidências de um ciclo previsível. Essa constatação reforça que a variação da *loss ratio* decorre de choques econômicos e estruturais, e não de um padrão cíclico regular. Assim, o indicador deve ser compreendido como um reflexo das condições de mercado e da eficiência operacional de cada companhia, sendo influenciado por fatores externos, como PIB, taxa de juros e eventos catastróficos.

Adicionalmente, o *Congressional Research Service* (2022) destaca que o aumento da volatilidade da sinistralidade, intensificado por fenômenos climáticos, crises econômicas e mudanças regulatórias, exige que as seguradoras adotem modelos de análise de cenários prospectivos. Esses modelos permitem estimar como diferentes condições futuras podem afetar a *loss ratio*, os prêmios e a solvência, favorecendo uma gestão de riscos mais preventiva e resiliente. Dessa forma, o monitoramento contínuo da razão sinistro/prêmio torna-se um instrumento estratégico não apenas para avaliar o desempenho passado, mas também para antecipar desequilíbrios técnicos e orientar políticas de precificação e capital.

Em síntese, a literatura indica que uma sinistralidade excessivamente alta está associada à perda de rentabilidade e à deterioração do equilíbrio atuarial, enquanto níveis muito baixos podem indicar super precificação ou retenção conservadora de riscos. A interpretação da taxa de sinistralidade em geral é definida pela própria seguradora, não há um intervalo universal, devendo a análise ser ajustada à natureza do produto, à estrutura de custos e às condições de mercado. Entretanto, o *Corporate Finance Institute* (2024) indica que valores da taxa de sinistralidade em torno de 40% a 60%, são geralmente considerados tecnicamente aceitáveis. Assim, a razão sinistro/prêmio consolida-se como uma métrica essencial para a avaliação do desempenho técnico das seguradoras, mas deve ser interpretada de forma contextual e integrada a outros indicadores de rentabilidade, solvência e risco.

2.4. Subscrição adversa, risco moral e equidade na precificação

Os mercados de seguros e planos de saúde são influenciados por problemas de informação assimétrica, os quais se manifestam principalmente por meio da seleção adversa e do risco moral. Esses eventos afetam o equilíbrio de mercado, a eficiência alocativa e a equidade na precificação, sendo amplamente estudados pela teoria microeconômica e pela economia da informação (NICHOLSON; SNYDER, 2010).

A seleção adversa, também conhecida como subscrição adversa, ocorre antes da contratação do seguro, quando o segurado possui informações privadas sobre seu risco individual que não são totalmente conhecidas pela seguradora. Como consequência, indivíduos com maior probabilidade de sinistro tendem a preferir planos mais abrangentes ou com prêmios mais baixos, enquanto aqueles de menor risco se afastam desses contratos. Esse processo aumenta o prêmio médio e reduz a participação de consumidores saudáveis, gerando o chamado efeito espiral de risco. (ROTHSCHILD; STIGLITZ, 1976).

Já o risco moral ocorre após a contratação, quando o comportamento do segurado muda em razão da cobertura obtida, levando-o a adotar condutas mais arriscadas ou a consumir serviços médicos de forma excessiva, por não internalizar integralmente os custos (PAULY, 1968).

Esses mecanismos são reforçados pelo desenho contratual dos planos. A teoria de contratos mostra que, diante da informação assimétrica, o principal (segurador) precisa desenhar incentivos adequados para o agente (segurado), como franquias, coparticipações e

limites de cobertura, de modo a mitigar o risco moral sem afastar clientes de baixo risco. Entretanto, tais mecanismos de eficiência podem comprometer a equidade no acesso e na precificação, ao penalizar indivíduos com maior propensão natural ao risco (NICHOLSON; SNYDER, 2010). Portanto, o dilema entre eficiência atuarial e justiça distributiva está no centro das discussões sobre políticas de precificação justa.

Estudos empíricos recentes aprofundam a distinção e a mensuração de cada um desses efeitos. Powell e Goldman (2021), em *Disentangling Moral Hazard and Adverse Selection in Private Health Insurance*, analisam dados administrativos de uma grande empresa norte-americana que alterou sua estrutura de planos de saúde entre 2005 e 2007. Utilizando modelos de escolha discreta e regressão quantílica generalizada, os autores conseguem separar os efeitos de risco moral e seleção adversa na distribuição de gastos médicos. Os resultados indicam que aproximadamente 47% do aumento dos custos médios nos planos mais generosos decorre de risco moral, enquanto 53% se devem à seleção adversa. Ou seja, ambos os fenômenos têm magnitudes semelhantes e exercem papel conjunto nas ineficiências do mercado.

O estudo também destaca que métodos simplificados, como avaliar a seleção adversa apenas com base nos gastos médicos do período anterior, tendem a subestimar sua real magnitude, devido à tendência natural de regressão à média (*mean reversion*). Além disso, os autores contestam a ideia de que os segurados respondem apenas ao preço marginal no fim do ano, demonstrando que o comportamento de consumo de saúde depende de todo o conjunto de incentivos não lineares do plano (dedutível, coparticipação e teto anual). Esses achados reforçam a importância de modelos contratuais e regulatórios mais flexíveis, capazes de captar a heterogeneidade dos comportamentos individuais.

De forma complementar, o *NBER Digest* (2016) sintetiza um estudo similar de Powell e Goldman, ressaltando que a distinção empírica entre seleção adversa e risco moral é essencial para políticas eficazes. O artigo mostra que as políticas públicas voltadas exclusivamente à redução de um desses efeitos podem agravar o outro. Por exemplo, o aumento da coparticipação reduz o risco moral, mas acentua a seleção adversa ao afastar indivíduos de alto risco; por outro lado, políticas de universalização do acesso, como o *Affordable Care Act* (ACA), mitigam a seleção adversa, mas podem elevar o consumo excessivo de serviços médicos. Assim, o equilíbrio entre eficiência econômica e equidade social depende da ponderação entre ambos os efeitos.

Nicholson e Snyder (2010) explicam que a solução ótima para esses problemas passa pelo desenho de contratos de seguro que combinem mecanismos de *screening*, que são

estratégias em que a seguradora oferece diferentes tipos de contratos para que os próprios consumidores revelem seu nível de risco por meio da escolha que fazem, (para lidar com seleção adversa) e incentivos comportamentais (para reduzir o risco moral). Em mercados competitivos, surgem dois tipos de equilíbrio: o *pooling equilibrium*, onde todos pagam o mesmo prêmio, promovendo equidade horizontal; e o *separating equilibrium*, onde os contratos se diferenciam conforme o tipo de risco, garantindo eficiência, mas reduzindo equidade. A precificação atuarial moderna busca justamente o ponto de equilíbrio entre essas duas dimensões, eficiência e justiça, especialmente sob a ótica regulatória e ética.

2.5. Modelagem Estatística em Seguros

A modelagem estatística desempenha um papel importante no âmbito atuarial e no processo de precificação de seguros, pois permite capturar padrões de risco, prever sinistros e segmentar carteiras de forma mais precisa. Métodos como os GLMs permitem a modelagem de variáveis de diferentes distribuições de probabilidade (como Poisson, Binomial e Gama), utilizando funções de ligação apropriadas, o que os torna amplamente aplicáveis na análise de dados de seguros para estimar a frequência de sinistros e a severidade das indenizações (MCCULLAGH; NELDER, 1989). Além disso, técnicas de regressão logística são frequentemente aplicadas para modelar a probabilidade de ocorrência de sinistro, tratando o desfecho como variável binária (ocorreu/não ocorreu), auxiliando na análise de fatores de risco e no processo de subscrição (HOSMER; LEMESHOW; STURDIVANT, 2013).

Já métodos de clusterização (como *k-means* ou análise hierárquica) permitem segmentar a carteira de clientes em grupos homogêneos de risco. O k-means, proposto por Lloyd (1957), agrupa dados em k clusters com base na minimização da distância entre pontos e seus centróides. Essa segmentação é particularmente útil em situações em que variáveis observadas não explicam totalmente os padrões de sinistralidade, contribuindo para identificar perfis de segurados e ajustar estratégias de precificação (JAIN, 2010). Portanto, a incorporação de técnicas estatísticas avançadas na modelagem atuarial amplia a capacidade da seguradora de compreender a estrutura de riscos da carteira, melhorar a acurácia das estimativas e implementar políticas tarifárias mais justas e eficientes.

3. METODOLOGIA

Neste capítulo, apresenta-se a metodologia adotada para análise da carteira de seguros. O objetivo é descrever os dados utilizados, os procedimentos de tratamento e integração das bases, bem como os métodos estatísticos empregados.

3.1. Base de dados

O presente estudo utiliza dados fornecidos por uma seguradora brasileira, nos ramos de seguros habitacional e de responsabilidade civil, que deseja se manter anônima. Tais dados abrangem o período de janeiro de 2015 a julho de 2025, correspondendo a 186.860 apólices diferentes e 7.501 endossos, totalizando 194.531 chaves apólice-endosso (cada apólice pode obter mais de 1 endosso). No período analisado, foram observados 21.821 sinistros.

As informações foram disponibilizadas em dois conjuntos principais:

1. Base de prêmios – Composta pelo relatório de Prêmios emitidos durante o mês de competência (PREMIT). Contendo dados de emissão de apólices, endossos, importância segurada, número de cadastro anonimizado (código apólice-endosso), UF, valores de prêmio total e líquido, ramo do seguro e demais características contratuais relevantes.
2. Base de sinistros – Formada pelo relatório de Sinistros Pagos (SINPAG) e Sinistros Pendentes (SINPEND). Trazendo registros de ocorrências indenizáveis, valor pago, valor reservado, ramo do seguro, número de cadastro anonimizado e número da apólice correspondente, número do sinistro, número do endosso, tipo de sinistro ou despesa, tipo de movimentação.

Para assegurar a confidencialidade das informações e atender às diretrizes de proteção de dados, todos os campos de identificação pessoal (como apólice, CPF/CNPJ e endereço) e campos de valores foram anonimizados, substituindo-se os identificadores originais por valores não rastreáveis. Cabe destacar que, embora anonimizados, os identificadores preservam a correspondência interna entre registros iguais, ou seja, valores idênticos no conjunto original foram substituídos por um mesmo identificador anônimo. Assim, é possível garantir a consistência das relações entre as variáveis sem comprometer a privacidade dos dados. Em acréscimo, para manter os dados de forma anonimizada, os valores de prêmios e de sinistros

foram multiplicados por duas constantes, para não se publicar dados reais da empresa. Porém nenhuma dessas transformações alteram os resultados aqui obtidos.

A Tabela *1* apresenta os ramos dos seguros que será trabalhado neste estudo, contendo código, sigla, nome e a sua descrição.

Tabela 1 - Ramos de seguros

Código do Ramo	**Sigla**	**Nome do Ramo**	**Descrição**
0313	RCA	Responsabilidade Civil Ambiental	Cobre danos ambientais causado pelo segurado a terceiros.
0351	RCG	Responsabilidade Civil Geral	Cobre danos a terceiros por fatos relacionados às atividades do segurado.
0378	RCP	Responsabilidade Civil Profissional	Cobre erros e omissões cometidos, por profissionais, no exercício da sua atividade.
1061	HAB MIP	Seguro Habitacional em Apólices de Mercado – Prestamista	Cobre o saldo devedor do financiamento habitacional quando o segurado falece ou sofre invalidez permanente.
1065	HAB DFI	Seguro Habitacional em Apólices de Mercado – Demais Coberturas	Cobre danos aos imóveis.
1068	HAB SFH	Seguro Habitacional Fora do Sistema Financeiro de Habitação (S.F.H.)	Cobre imóveis que não fazem parte do Sistema Financeiro da Habitação. Cobre danos físicos ao imóvel, Responsabilidade civil ou riscos associados ao uso da propriedade.

Fonte: A autora (2025)

3.2. Processamento e organização dos dados

Inicialmente, as variáveis passaram por um processo de limpeza e padronização, corrigindo inconsistências de formatação, padronizando datas, uniformizando variáveis categóricas e tratando valores nulos ou inconsistentes. Em seguida, foi realizada a integração entre as bases de prêmios e sinistros por meio do número de identificador seguro, permitindo a construção de um histórico consolidado de cada contrato, com seus respectivos valores de prêmio e sinistros pagos.

As principais etapas de processamento incluíram: 1) Eliminação de registros duplicados e inconsistentes; 2) conversão de variáveis categóricas para formato padronizado; 3) cálculo da sinistralidade individual por apólice; 4) criação de variáveis derivadas, como frequência de sinistros e severidade média. Para saneamento dos dados foi utilizado o Microsoft Excel, no início tinha-se 3 bases, 2 de sinistros e 1 de prêmio, totalizando 230.954 e 3.124 entradas de prêmio e sinistros respectivamente, após saneamento unificamos as bases, o que resultou em 199.286 linhas.

3.3. Análise de Cluster

Clustering é uma técnica destinada a identificar agrupamentos naturais dentro de um conjunto de dados, reunindo objetos que apresentam padrões semelhantes quando avaliados por múltiplas características. De maneira operacional, esse processo consiste em particionar um conjunto de n elementos em k grupos com base em medidas de similaridade, de modo que os itens pertencentes ao mesmo grupo sejam mais próximos entre si do que em relação aos demais.

Na prática, esses agrupamentos podem variar quanto à forma, tamanho e densidade, e a presença de ruído torna a identificação das estruturas ainda mais desafiadora. Embora um *cluster* ideal seja descrito como um conjunto compacto e bem separado, sua interpretação é subjetiva e depende do conhecimento do domínio. Essas particularidades, combinadas com o aumento da dimensionalidade dos dados, explicam o desenvolvimento contínuo de inúmeros algoritmos de agrupamento ao longo das últimas décadas (JAIN, 2010).

3.4. Regressão Logística

A regressão logística é um modelo estatístico utilizado para analisar relações entre uma variável dependente binária e um conjunto de variáveis explicativas. Conforme apresentado por Hosmer, Lemeshow e Sturdivant (2013), o modelo descreve a probabilidade condicional do evento de interesse por meio da função logística, garantindo que os valores previstos permaneçam entre 0 e 1 e permitindo interpretar os efeitos covariados no logito.

A escolha desse método é diretamente condicionada às características dos dados analisados, sobretudo ao tipo de variável dependente. Quando o desfecho assume apenas dois valores possíveis (ocorrência ou não do evento), a regressão linear se torna inadequada, pois pode produzir valores previstos fora do intervalo entre 0 e 1 e viola pressupostos básicos de homoscedasticidade.

Nesse contexto, a regressão logística se apresenta como o método mais apropriado, uma vez que modela a probabilidade do evento por meio da função logística, garantindo valores compatíveis com interpretações probabilísticas. Além disso, esse modelo acomoda naturalmente variáveis explicativas de diferentes naturezas, como contínuas, categóricas e ordinais, permitindo incorporar informações relevantes sem comprometer a estrutura do modelo. Outros aspectos dos dados também influenciam a escolha e o desempenho da técnica, como a presença de colinearidade entre covariáveis, o número de eventos observados e o desenho amostral.

3.5. Estratégia de análise

A análise foi estruturada em quatro etapas complementares, refletindo a complexidade da carteira e os diferentes objetivos do estudo. A primeira etapa consistiu em uma análise descritiva das principais variáveis atuariais da base de dados, com medidas de tendência central, dispersão e visualizações gráficas por ramo e por UF. Essa fase teve como foco caracterizar a distribuição dos prêmios, sinistros e sinistralidade, além de identificar padrões assimétricos e concentrações de risco relevantes para as análises subsequentes.

A segunda etapa concentrou-se na análise da eficiência técnica da carteira por meio da sinistralidade, abordando sua distribuição empírica geral e segmentada por ramo, incluindo histogramas, *violin plots* e *boxplots*. Essa etapa também envolveu o cálculo da razão

sinistro/prêmio, da frequência e da severidade média dos sinistros, com ênfase na heterogeneidade entre os contratos.

A terceira etapa envolveu a segmentação da carteira por meio de análise de cluster. Foram utilizadas variáveis técnico-atuariais como prêmio, sinistro e frequência. Os dados foram padronizados (z-score) e agrupados com o método k-means, com definição do número de grupos via métodos do cotovelo e da silhueta (Rousseeuw, 1987; Jain, 2010). A segmentação visou identificar padrões consistentes de subscrição adversa, risco moral e heterogeneidade técnica entre as apólices.

A quarta e última etapa consistiu na modelagem estatística da probabilidade de ocorrência de sinistro, utilizando regressão logística binária com o prêmio como variável contínua e o ramo e a UF como efeitos fixos. A performance preditiva foi avaliada por meio da curva ROC (Característica de Operação do Receptor) e do AUC (Área Sob a Curva), conforme abordado por Hosmer, Lemeshow e Sturdivant (2013). Complementarmente, foi estimada a frequência condicional dos sinistros entre as apólices sinistradas, e testados modelos de contagem (Binomial Negativa) para avaliar a adequação estatística à distribuição observada.

Para todas as análises aqui apresentadas foi utilizado o R (pacotes WDI, dplyr, readxl, writexl, lubridate, tidyr, scales, ggh4x, FSA, tidyverse, rstatix, MASS, factoextra, ggplot2, broom, pROC e cluster).

3.6. Dados

A construção inicial da base de dados teve como etapa fundamental a definição da unidade de análise que seria utilizada ao longo de todo o trabalho. Como informado anteriormente, os arquivos fornecidos pela seguradora continham informações de prêmios, sinistros pagos e sinistros pendentes de pagamento, registrados a partir de diferentes tipos de movimentos (emissões, endossos e registros de sinistros). Em razão desse formato operacional, uma mesma apólice podia aparecer diversas vezes na base, o que exigia uma padronização capaz de evitar duplicidades e assegurar consistência na mensuração da exposição ao risco.

Nesse contexto, definiu-se que a unidade de análise seria a combinação apólice–endosso, construída pela chave composta por *NUM_APOL* e *NUM_END*. Essa escolha metodológica é coerente com a prática atuarial, uma vez que cada endosso representa uma

alteração contratual válida e pode estar vinculado a prêmios adicionais, ajustes de cobertura ou à ocorrência de sinistros específicos. Dessa forma, a consolidação por apólice-endosso preserva a granularidade necessária para análises de risco, evita superestimações do volume segurado e proporciona uma visão mais fiel das frequências e dos valores indenizados.

Após consolidada a base, obteve-se um conjunto total de 199.286 linhas, sendo 194.531 correspondentes a apólice-endosso válidas dentro do período estudado. A partir dessa estrutura foi possível identificar quantas apólices não tiveram nenhum sinistro, quantas registraram um único evento e quantas apresentaram múltiplos sinistros ao longo do período.

Para o tratamento dos prêmios, foram feitas, através do Excel, a movimentação de emissão (-) cancelamento para cada apólice, obtendo-se o valor de prêmio líquido. No caso dos sinistros, foi considerado os sinistros pagos e pendentes, subtraindo os salvados e ressarcidos, para obter o montante líquido total pago para a apólice. O cálculo da sinistralidade foi feito conforme a Equação 1.

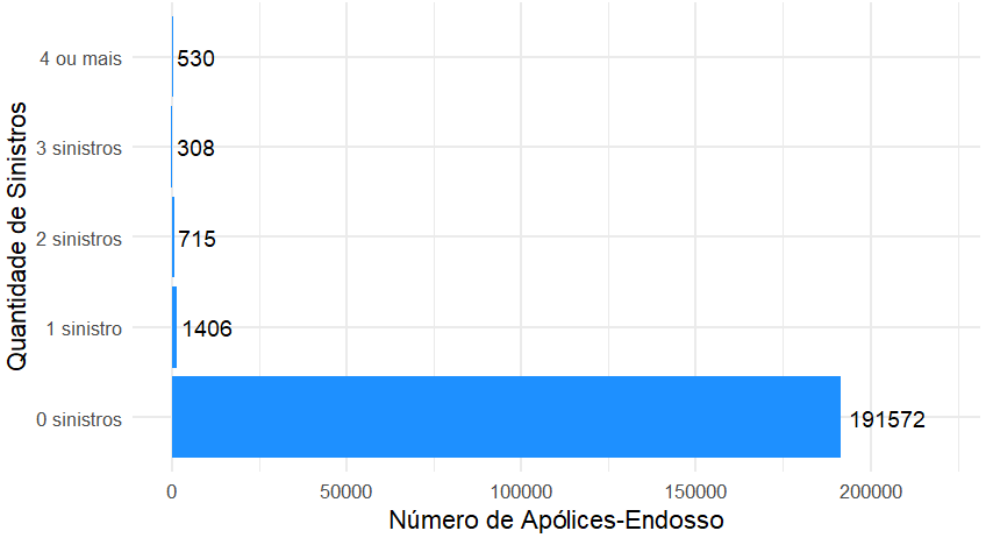
4. ANÁLISE DE DADOS E RESULTADOS

Como afirmado anteriormente, a sinistralidade representa a relação entre o total de sinistros ocorridos (pagos e pendentes) e o total de prêmios líquidos emitidos no período, constituindo um dos principais indicadores de equilíbrio técnico e desempenho atuarial das carteiras de seguros, permitindo verificar em que medida o volume de prêmios arrecadado é suficiente para cobrir os custos associados aos sinistros.

4.1. Análise Descritiva

Observou-se que a maioria das apólices-endosso não registrou sinistro algum, enquanto uma parcela relativamente pequena concentrou todos os eventos registrados. Das apólices analisadas, 21.821 sinistros diferentes foram observados, 1.406 apresentam pelo menos 1 sinistro, 715 registraram dois sinistros, 308 tiveram três sinistros e 530 apresentaram quatro ou mais ocorrências. Essa distribuição reflete o comportamento esperado em carteiras de ramos não vida: muitos segurados com poucos ou nenhum sinistro e poucos segurados com múltiplos eventos, que pode ser explicado principalmente no grupo habitacional, em que dispões de contratos coletivos e de vigências maiores.

Figura 1 - Distribuição de Apólice-Endosso por quantidade de sinistros

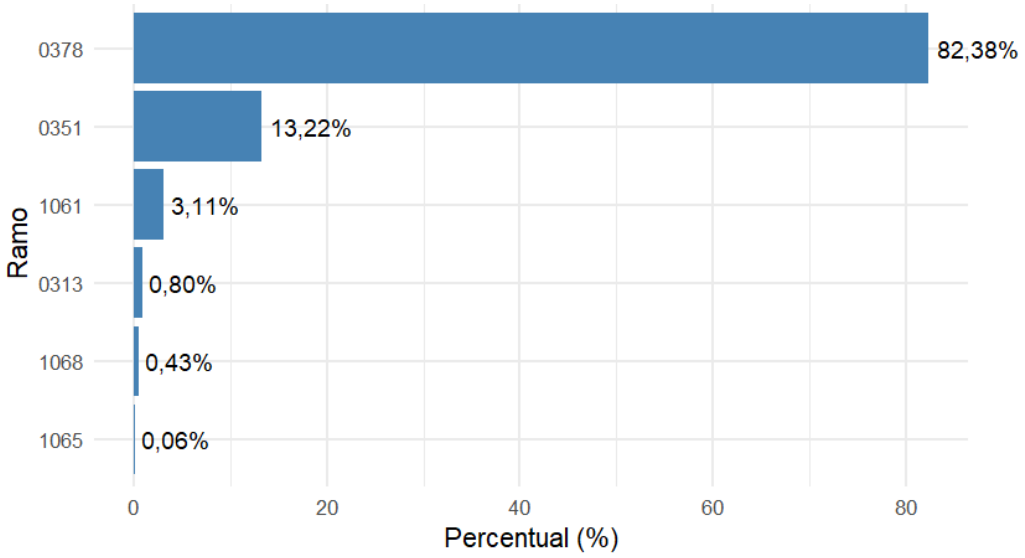


Fonte: A autora (2025)

Na sequência, procedeu-se à distribuição das apólices-endosso e dos sinistros por ramo de seguro, com o objetivo de identificar a participação de cada ramo tanto na exposição quanto no risco efetivamente realizado. Esse cruzamento mostrou-se particularmente relevante porque distintos ramos possuem perfis de risco naturalmente diversos, o que influencia diretamente a sinistralidade esperada, o comportamento da frequência e a necessidade de segmentação nos modelos preditivos.

Os dados mostram que o ramo RCP concentra a maior parte das apólices da carteira, representando cerca de 82% do total analisado. No entanto, quando olhamos para os sinistros, sua participação cai para aproximadamente 18%, indicando que, apesar de ser o ramo mais expressivo em volume, não é o que apresenta maior frequência de ocorrências.

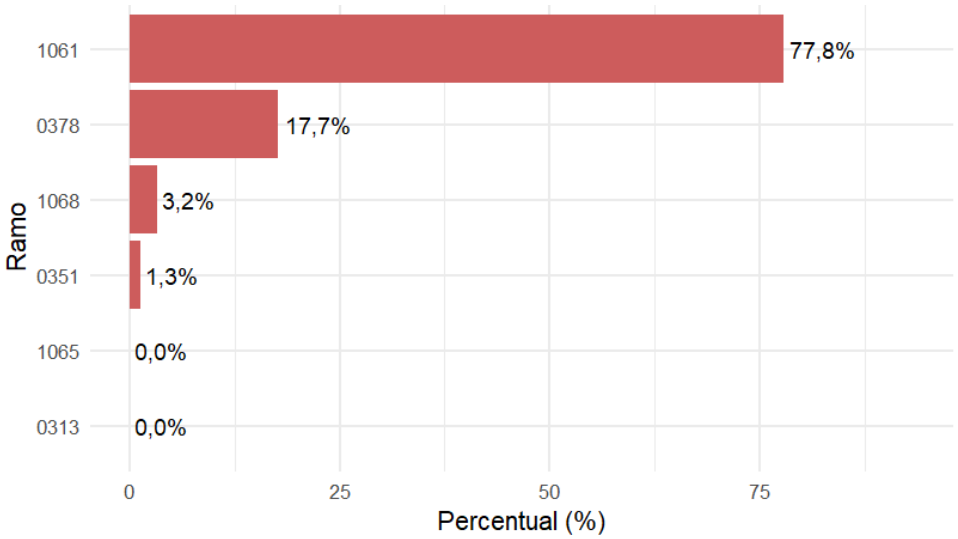
Figura 2 – Participação dos ramos no total de Apólice-Endosso



Fonte: A autora (2025)

Por outro lado, o ramo HAB MIP (1061) apresenta um cenário oposto: embora responda por apenas 3% das apólices, concentra cerca de 78% de todos os sinistros registrados. Esse contraste evidencia diferenças estruturais importantes entre os produtos e mostra que a carteira possui áreas com níveis de risco bastante distintos. Por isso, faz sentido aprofundar a análise de forma segmentada nas próximas etapas do estudo.

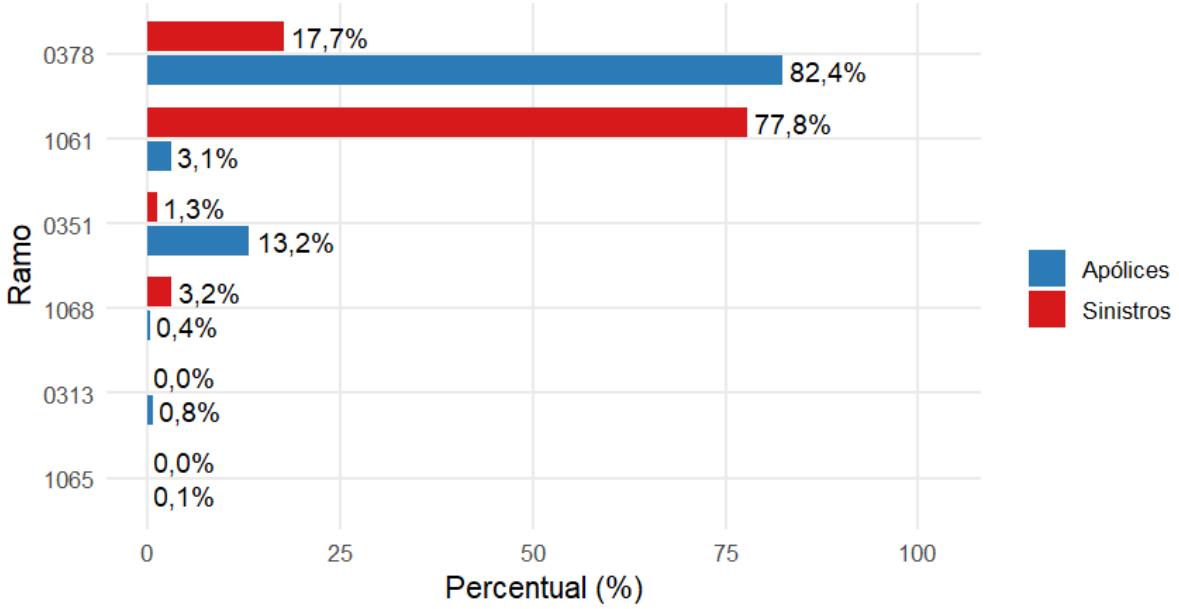
Figura 3- Participação dos ramos no total de Sinistros



Fonte: A autora (2025)

Para ilustrar essa relação entre exposição e risco, elaborou-se um gráfico em barras horizontais que compara, lado a lado, o percentual de apólices e o percentual de sinistros atribuídos a cada ramo. Esse gráfico pode ser observado na Figura *4*, pois ele sintetiza visualmente a discrepância entre a distribuição da carteira e a distribuição dos eventos de sinistro. O ramo RCP (0378) tem um grande volume de apólices, mas baixa participação nos sinistros, enquanto o ramo HAB MIP (1061) apresenta o comportamento oposto: concentra a maior parte dos eventos mesmo com poucas exposições. Isso mostra que contar apólices não é suficiente para entender o risco da carteira. É preciso considerar variáveis que representem melhor sua estrutura e dinâmica, como frequência e severidade dos sinistros. Essa visão é o ponto de partida para aplicar técnicas mais avançadas, como modelos de clusterização, análises de equidade e regressão logística, que ajudam a estimar a probabilidade de ocorrência de sinistros de forma mais precisa.

Figura 4 - Participação por ramo: Apólices x Sinistros



Fonte: A autora (2025)

A etapa de caracterização da base vai além de simplesmente descrever os dados. Ela cria o contexto essencial para todas as análises que virão. Esse processo revela a forte concentração dos sinistros em poucos ramos, evidencia a heterogeneidade entre eles e reforça a necessidade de segmentar e modelar a carteira considerando as particularidades de cada grupo. Ao consolidar essas informações iniciais, foi estabelecida uma estrutura sólida para avançar com análises mais complexas, como subscrição adversa, equidade na precificação e modelagem da probabilidade de sinistros, garantindo consistência metodológica e rigor estatístico em todas as etapas do trabalho.

Inicialmente foi feita uma análise exploratória de dados, através do R, desconsiderando os valores negativos. A Tabela *2* apresenta as medidas descritivas das principais variáveis da carteira analisada, incluindo prêmios líquidos emitidos, total de sinistros pagos e a sinistralidade.

Tabela 2 - Estatísticas descritivas da carteira de seguros total

Variável	Min	1º Quartil	Mediana	Média	3º Quartil	Máximo
Prêmio Emitido	1	929	2.090,00	29.821,00	5.371,00	48.481.786,00
Total Sinistro	-	-	-	5.670,00	-	21.739.047,00
Sinistralidade	-	-	-	0,15	-	492,58

Fonte: A autora (2025)

Observa-se que os prêmios emitidos apresentam alta dispersão, variam de aproximadamente R$1,00 (valor mínimo) até cerca de R$ 48 milhões (valor máximo). A média de aproximadamente R$ 29,8 mil é superior à mediana (R$ 2,09 mil), indicando forte assimetria à direita, ou seja, há poucas apólices de grande porte que elevam a média geral.

Em relação aos sinistros pagos, observa-se valores que vão aproximadamente até R$ 21,7 milhões, com média em torno de R$ 5,6 mil. A discrepância entre média e mediana também evidencia grande heterogeneidade entre as apólices, reflexo de sinistros isolados de alto impacto.

A sinistralidade apresenta mediana de 0 e média de 0,15, com valor máximo acima de 492, reforçando a presença de casos extremos em que os sinistros superaram significativamente os prêmios arrecadados. Essa alta variabilidade indica a existência de grupos de risco heterogêneos, nos quais determinados segmentos apresentam equilíbrio técnico, enquanto outros operam com forte déficit técnico. Esse comportamento confirma a natureza altamente assimétrica e concentrada da sinistralidade no setor securitário: a maioria das apólices apresenta sinistros nulos ou moderados, enquanto uma pequena parcela gera perdas expressivas, responsáveis por grande parte do custo total da carteira.

Para entender melhor a distribuição desses valores, separamos a análise também por grupo de ramo. A Tabela *3* apresenta o resumo estatístico separado para o grupo Habitacional, que obtém valores expressivos de prêmios e sinistros, com média de R$ 468 mil em prêmios emitidos e R$ 90 mil em sinistros totais. Contudo, há forte assimetria positiva, já que as medianas (R$ 54mil e R$ 0, respectivamente) são muito inferiores às médias. Os valores máximos evidenciam a presença de alguns contratos de grande porte que elevam substancialmente a média da amostra.

A sinistralidade média de 0,23 sugere que, em termos agregados, o ramo apresenta equilíbrio técnico satisfatório, com os sinistros correspondendo a cerca de 23% dos prêmios arrecadados. A mediana ainda menor reforça que a maioria das apólices opera com superávit técnico, e apenas uma pequena parcela concentra sinistros mais elevados. Esse comportamento é consistente com carteiras habitacionais, geralmente de baixa frequência e alta severidade, impactadas por eventos pontuais de maior magnitude.

Por fim, a Tabela *4* apresenta o ramo de responsabilidade civil mostrando uma estrutura muito mais dispersa e volátil. Embora os prêmios sejam menores que os observados no ramo habitacional (média de R$ 3,4 mil), a sinistralidade média (0,14) e o valor máximo evidenciam graves desequilíbrios técnicos em parte da carteira. A diferença entre média e mediana (6,11)

indica que a maioria das apólices opera com resultados moderados, mas há sinistros extremos que elevam consideravelmente o indicador agregado.

Tabela 3- Estatísticas descritivas da carteira de seguros Habitacionais

Variável	Min	1º Quartil	Mediana	Média	3º Quartil	Máximo
Prêmio Emitido	20,25	15.148,86	54.451,48	468.577,71	151.041,17	48.481.786,28
Total Sinistro	-	-	-	90.703,69	-	21.739.047,43
Sinistralidade	-	-	-	0,23	-	430,65

Fonte: A autora (2025)

Tabela 4- Estatísticas descritivas da carteira de seguros de Responsabilidade civil

Variável	Min	1º Quartil	Mediana	Média	3º Quartil	Máximo
Prêmio Emitido	0,87	811,52	1.979,28	3.372,98	4.659,09	307.137,64
Total Sinistro	-	-	-	544,51	-	3.293.060,78
Sinistralidade	-	-	-	0,14	-	492,58

Fonte: A autora (2025)

De modo geral, nota-se que o ramo Habitacional apresenta os maiores volumes de prêmio emitido e sinistros, além de uma ampla dispersão nos valores, evidenciada pela grande diferença entre média e mediana. Apesar dessa oscilação, a sinistralidade média (0,23) é relativamente baixa, o que sugere maior eficiência técnica e desempenho estável, ainda que concentrado em poucas apólices de grande valor.

Em contrapartida, o ramo de Responsabilidade Civil exibe prêmios e sinistros de menor magnitude, porém com alta variabilidade relativa na sinistralidade (média de 0,14 e máximo de 492,58). Esse comportamento reflete uma maior exposição a risco técnico, típica de ramos em que os sinistros são menos frequentes, mas potencialmente mais severos.

Foram elaboradas visualizações descritivas para explorar o comportamento das variáveis técnicas da carteira, incluindo prêmios, sinistros e sinistralidade. Primeiramente, foi realizada uma análise das variáveis prêmio líquido emitido e sinistro total pago, com o objetivo de comparar visualmente a dispersão e a relação entre os valores arrecadados e os valores indenizados pela seguradora. A fim de reduzir o impacto de *outliers* que poderiam distorcer a escala do gráfico, e facilitar a visualização e análise do mesmo, limita-se o intervalo de valor até R$ 100 mil e, considerando apenas valores maiores que zero, além disso, os valores foram convertidos para milhares de reais (R$ mil).

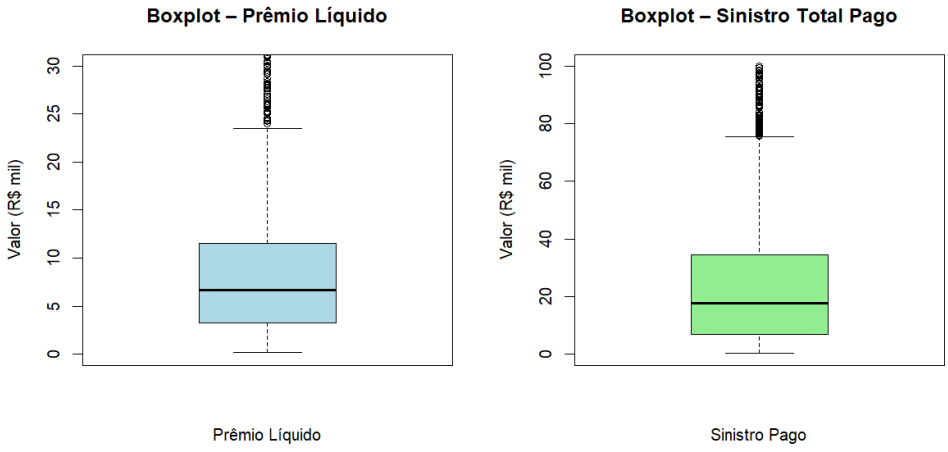
Cada variável foi representada por uma cor distinta, azul para prêmio líquido e verde para sinistro total pago. Apesar dos dados estarem limitados a R$100 mil, a escala do eixo vertical foi limitada a R$30 mil reais para os prêmios e R$ 100 para os sinistros, a fim de ampliar a visualização da região central das distribuições.

A Figura *5* permite observar a distribuição e a dispersão dos valores financeiros relacionados aos prêmios recebidos e aos pagamentos dos sinistros. Percebe-se que o Prêmio Líquido possui distribuição mais concentrada, com mediana próxima de R$ 6 mil e poucos valores acima de R$ 20 mil, embora existam alguns *outliers* positivos, indicados pelos círculos acima do limite superior. Essa concentração sugere que a maior parte das apólices possui valores de prêmio relativamente baixos, enquanto poucas apólices apresentam prêmios expressivamente elevados, o que é típico em carteiras com contratos de diferentes portes.

Já o Sinistro Total Pago apresenta maior dispersão e amplitude, com mediana próxima a R$ 15 mil e valores máximos superiores a R$ 80 mil, além de numerosos registros acima desse patamar. Isso indica que, embora a maioria das apólices tenha sinistros de baixo valor (ou nenhum sinistro), há eventos pontuais de alta severidade, que elevam o custo médio e impactam a sinistralidade.

De forma geral, a diferença entre as amplitudes dos dois boxplots revela que a variabilidade dos sinistros é significativamente maior do que a dos prêmios, refletindo a natureza incerta e concentrada das perdas no seguro. Esse comportamento é esperado no contexto atuarial, pois enquanto os prêmios tendem a ser definidos de forma previsível e estável, os sinistros são eventos aleatórios e de magnitude variável.

Figura 5 - Boxplot Prêmio Líquido e Sinistro



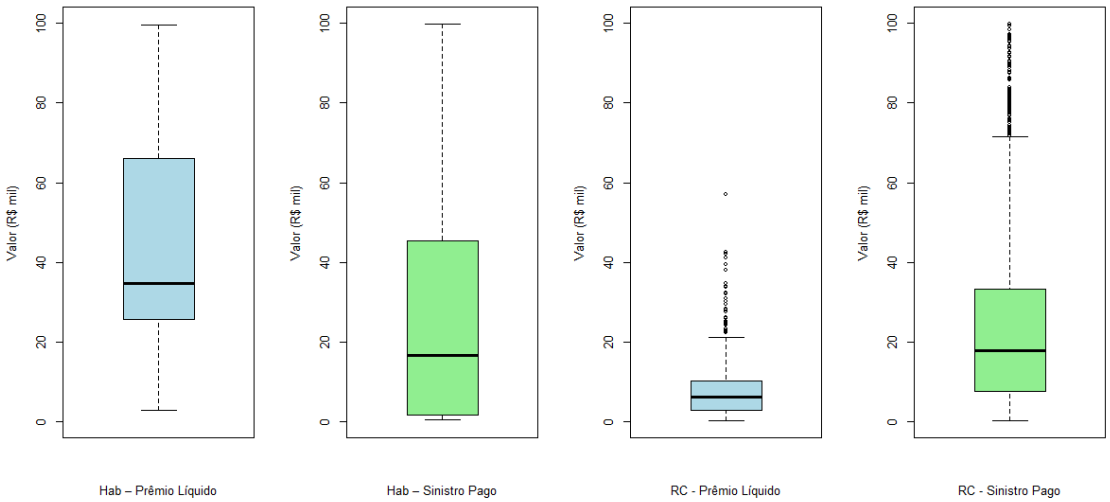
Fonte: A autora (2025)

Para aprofundar a análise descritiva da eficiência técnica, foi elaborado um boxplot comparativo por grupo de ramo, considerando os mesmos padrões do boxplot completo, porém com escala limitada a R$100 mil, com o objetivo de verificar o comportamento e impacto de cada ramo em meio a base de dados geral analisada.

A Figura *6* apresenta a distribuição dos valores de Prêmio Líquido e Sinistro Pago para os ramos Habitacional e Responsabilidade Civil. O Habitacional apresenta valores médios e medianos mais elevados, tanto em prêmios quanto em sinistros, além de maior amplitude total. Isso confirma que esse ramo representa maior volume financeiro, refletindo contratos de maior porte e prêmios proporcionalmente mais altos. Já o ramo Responsabilidade Civil exibe valores mais reduzido de prêmio, porém com alta dispersão nos sinistros pagos, incluindo uma quantidade considerável de outliers positivos (sinistros de alta severidade). Essa característica é típica de carteiras em que eventos de indenização são menos frequentes, mas potencialmente mais custosos, o que eleva o risco técnico.

Em ambos os ramos, a mediana do sinistro pago é inferior à do prêmio líquido, o que sugere desempenho técnico positivo (superávit). No entanto, a presença de valores extremos indica que poucas apólices específicas concentram perdas significativas, podendo distorcer a média e afetar a sinistralidade global.

Figura 6- Boxplot distribuição de prêmio líquido e Sinistro por ramo

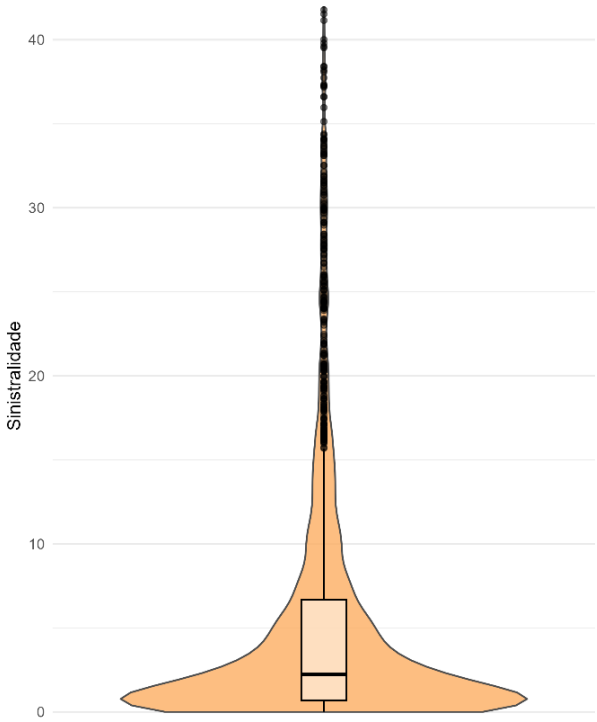


Fonte: A autora (2025)

O *violin plot* da sinistralidade, Figura *7*, foi elaborado com o objetivo de visualizar sua distribuição completa em toda a base de apólices, destacando não apenas as medidas de tendência central e dispersão, mas também a densidade de concentração dos valores. O formato afunilado na parte superior e mais largo na base indica que a maior parte das apólices possui sinistralidade baixa, concentrando-se em valores próximos a zero, enquanto uma fração menor de contratos apresenta índices significativamente mais elevados.

O boxplot interno reforça essa leitura: a mediana da sinistralidade situa-se em torno de valores baixos, denotando que a maioria das apólices se mantém dentro de um desempenho técnico favorável (superávit). No entanto, a amplitude interquartílica relativamente grande demonstra variabilidade considerável entre os contratos, especialmente entre o terceiro quartil e os valores máximos.

Figura 7- Violin Plot de distribuição de sinistralidade



Fonte: A autora (2025)

A Figura *8* apresenta a distribuição da sinistralidade segmentada por grupo de ramo, distinguindo os seguros Habitacional e de Responsabilidade Civil. Essa divisão evidencia padrões de comportamento distintos entre as duas carteiras. A construção do gráfico seguiu os

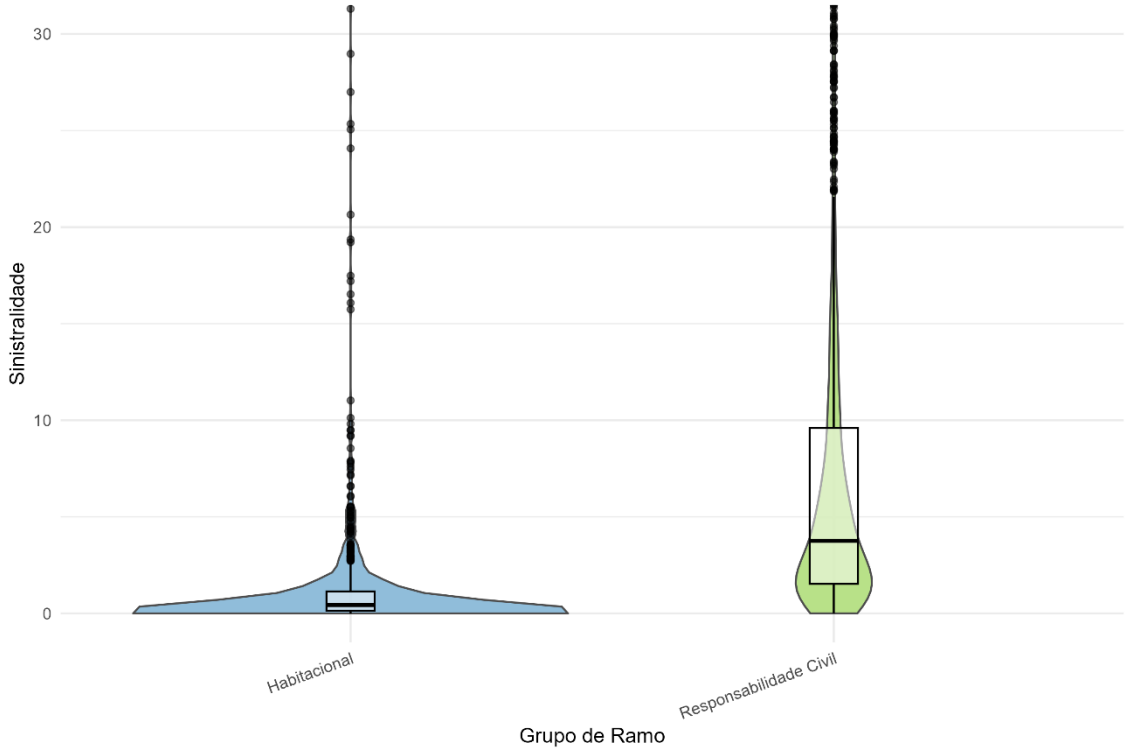
mesmos critérios adotados no *violin plot* geral, mas restringiu a sinistralidade ao intervalo entre 0 e 30, a fim de eliminar valores extremos e tornar a visualização mais informativa.

O ramo Habitacional apresenta uma distribuição concentrada em torno de valores baixos, com mediana próxima de zero, indicando que a maior parte das apólices possui baixo índice de sinistros, enquanto uma minoria apresenta eventos de alta severidade.

Em contraste, o ramo de Responsabilidade Civil apresenta maior dispersão vertical e maior densidade em faixas intermediárias de sinistralidade, o que indica uma sinistralidade mais heterogênea e volátil. Além disso, nota-se que a mediana do RC (Responsabilidade civil) é mais alta que a do Habitacional, o que sugere eficiência técnica inferior nesse grupo, isto é, proporcionalmente, as indenizações consomem parcela maior dos prêmios arrecadados.

De modo geral, o formato mais "achatado" do violino no ramo Habitacional revela concentração em baixos níveis de sinistralidade e predominância de apólices superavitárias, enquanto o formato mais "estreito e alongado" no RC reforça alta variabilidade técnica e maior risco agregado.

Figura 8 - Violin Plot da Distribuição da Sinistralidade por Grupo de Ramo



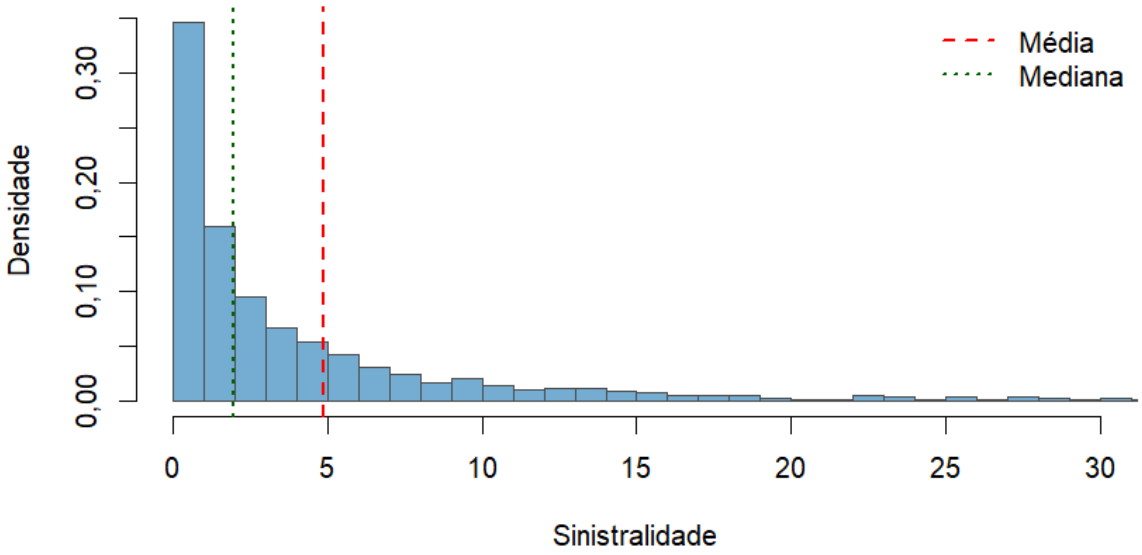
Fonte: A autora (2025)

Para complementar a análise descritiva da sinistralidade, foi construído um *True Histogram* (Figura *9*), que permite visualizar de forma detalhada a distribuição empírica da Sinistralidade. Esse tipo de gráfico difere do histograma convencional por utilizar larguras de classe (*bins*) uniformes e representar a densidade de probabilidade em vez da contagem bruta de observações.

No caso em análise, a variável sinistralidade foi limitada ao intervalo de 0 a 30, de modo a excluir valores extremos e focar na faixa que representa o comportamento técnico típico da carteira. A linha vermelha tracejada indica a média da sinistralidade, enquanto a linha verde pontilhada representa a mediana, permitindo comparar visualmente a posição e a assimetria da distribuição.

O histograma evidencia uma distribuição fortemente assimétrica à direita, com elevada concentração de observações próximas a zero, o que indica que a maior parte das apólices apresenta baixa sinistralidade ou sequer registrou sinistro no período analisado. Essa predominância de valores baixos revela características de boa eficiência técnica e controle de risco na maior parte da carteira. A cauda longa à direita demonstra a presença de eventos de alta severidade, típicos em carteiras de seguros, visto que poucos contratos concentram grandes perdas, o que impacta fortemente o resultado técnico agregado.

Figura 9- True Histogram da distribuição da sinistralidade



Fonte: A autora (2025)

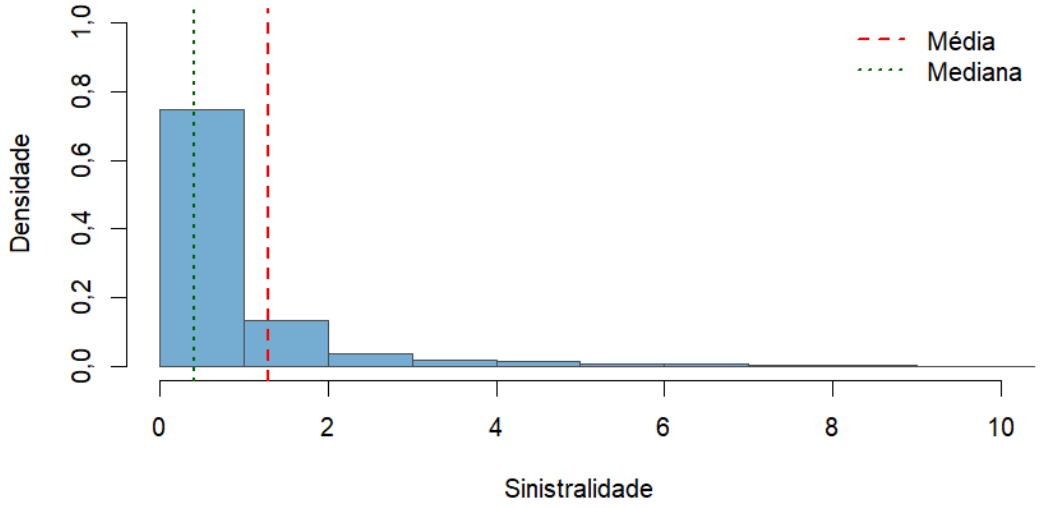
No *True Histogram* construido para o ramo Habitacional, Figura *10*, a distribuição da sinistralidade mantém o padrão de forte assimetria positiva observado na carteira como um

todo. Há elevada concentração de contratos com sinistralidade muito baixa, enquanto um número reduzido de observações apresenta sinistralidades elevadas, prolongando a cauda à direita até valores próximos de 10. Esse comportamento evidencia uma cauda longa, embora menos extrema em comparação à distribuição geral, indicando que poucos eventos de alta severidade ainda exercem influência significativa sobre a média e sobre a avaliação do desempenho técnico do ramo.

Por outro lado, o Ramo de Responsabilidade Civil (

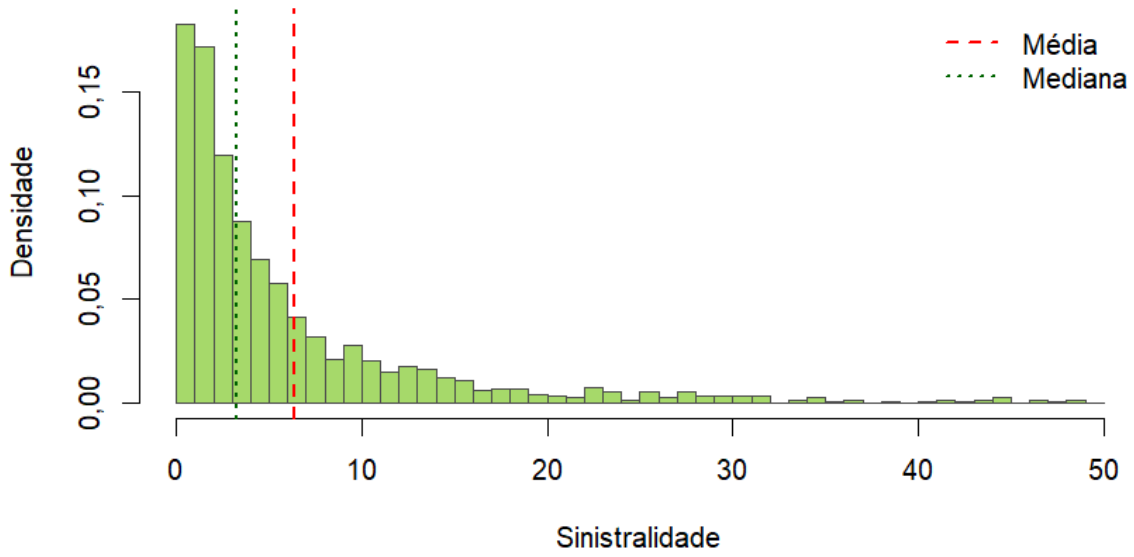
Figura 11) apresenta uma distribuição mais dispersa e alongada, com média e mediana mais elevadas e uma cauda também longa à direita. Essa configuração revela maior variabilidade e risco técnico, indicando que, embora a maior parte das apólices ainda mantenha sinistralidade reduzida, há uma quantidade mais relevante de casos com índices de sinistralidade acima da média. Esse comportamento é típico de ramos com baixa frequência e alta severidade, em que poucos sinistros podem gerar perdas expressivas.

Figura 10 - Histograma da sinistralidade para o ramo Habitacional



Fonte: A autora (2025)

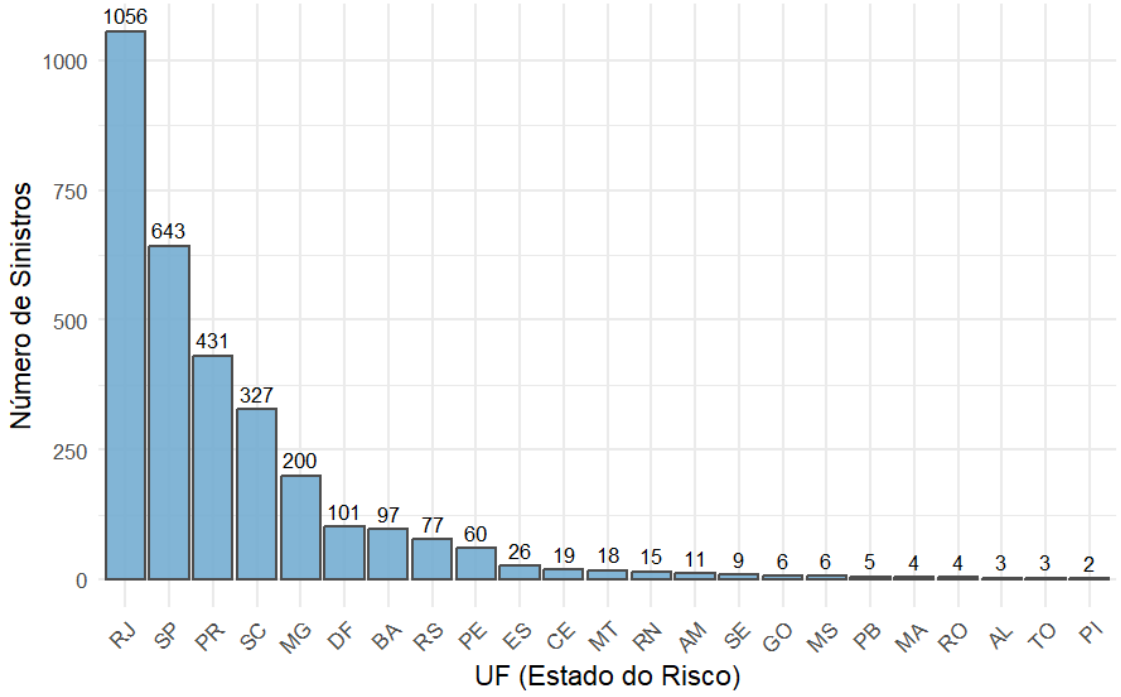
Figura 11– Histograma da sinistralidade para o ramo de Responsabilidade Civil



Fonte: A autora (2025)

A diferença entre as posições das linhas da média (vermelha) e da mediana (verde) é mais pronunciada no ramo de Responsabilidade Civil, evidenciando a influência de *outliers* e eventos extremos sobre o resultado agregado. No Habitacional, essa diferença é menor, indicando maior homogeneidade e estabilidade técnica.

Figura 12- Histograma da quantidade de sinistros por UF



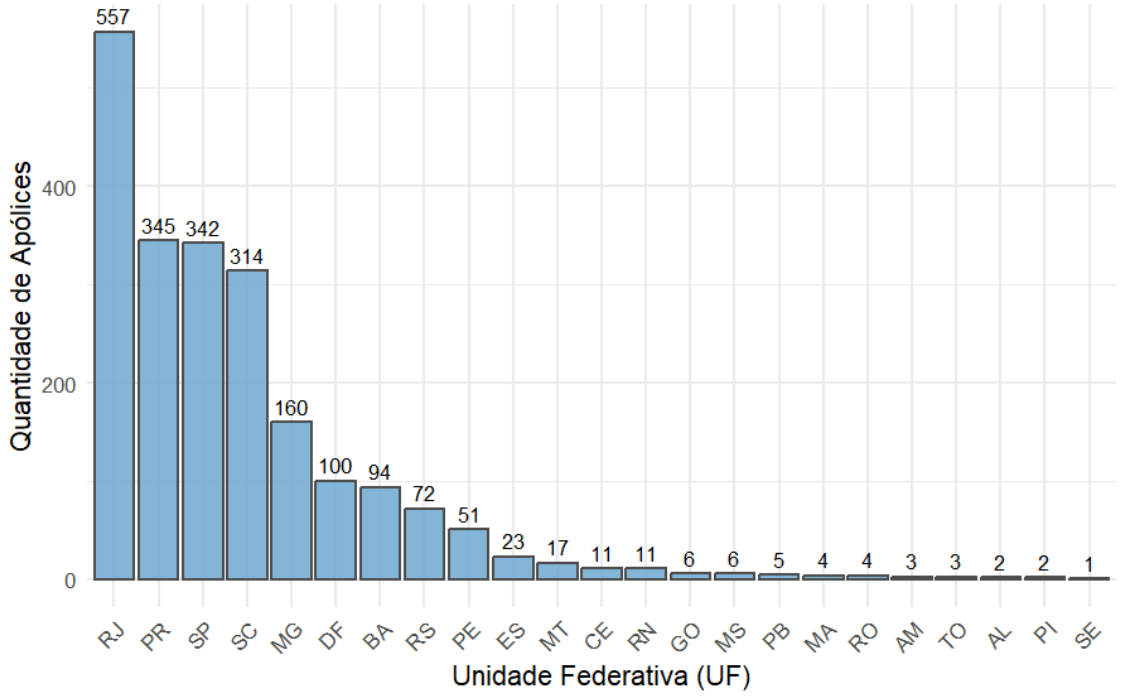
Fonte: A autora (2025)

O conjunto de histogramas por UF, Figura 12, revela que o Rio de Janeiro e São Paulo concentram tanto o volume de apólices quanto a ocorrência de sinistros, sendo responsáveis pela maior parcela do resultado técnico da carteira.

O Ramo de Responsabilidade Civil apresenta maior exposição e volatilidade técnica, refletindo risco elevado, enquanto o Habitacional demonstra perfil mais estável e previsível, com baixa sinistralidade e concentração moderada. Essa configuração reforça a importância de estratégias de diversificação geográfica e ajuste fino na precificação, especialmente para o ramo de Responsabilidade Civil.

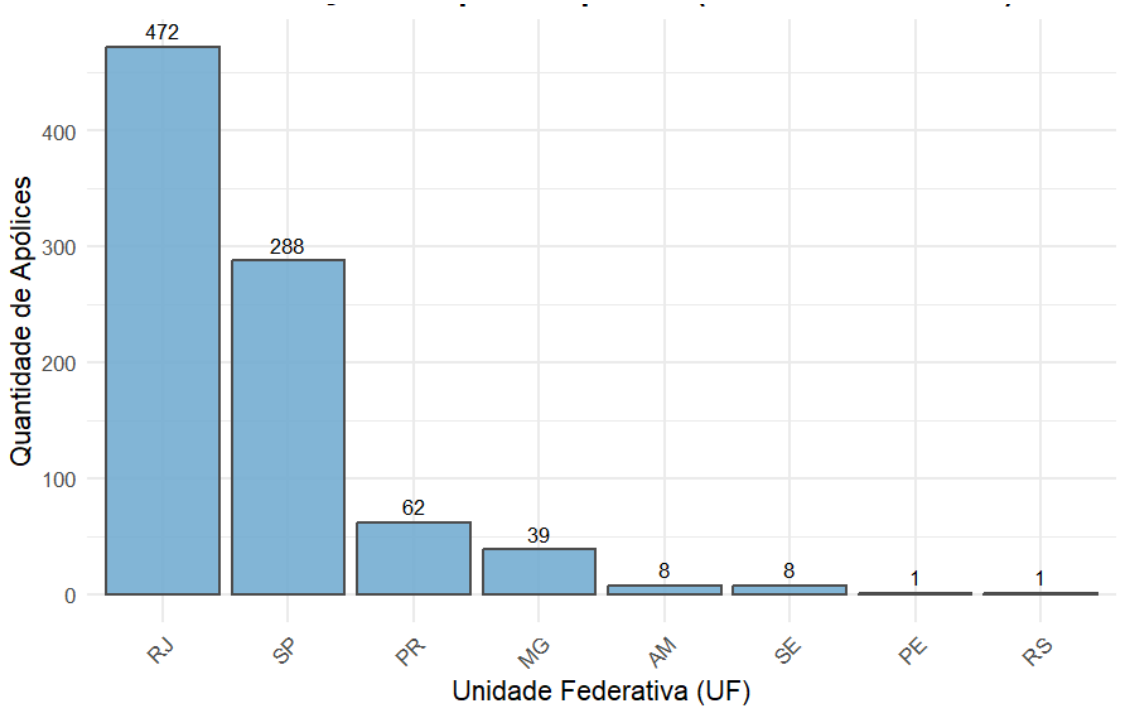
A Figura 13 apresenta o número de apólices por UF para ramo de responsabilidade civil. A Figura 14 apresenta a quantidade de apólices por UF para o ramo Habitacional. Enquanto que as Figura 15 e Figura *16* apresentam a quantidade de sinistros para os ramos habitacional e de responsabilidade civil, respectivamente.

Figura 13- Distribuição de apólices do Responsabilidade Civil



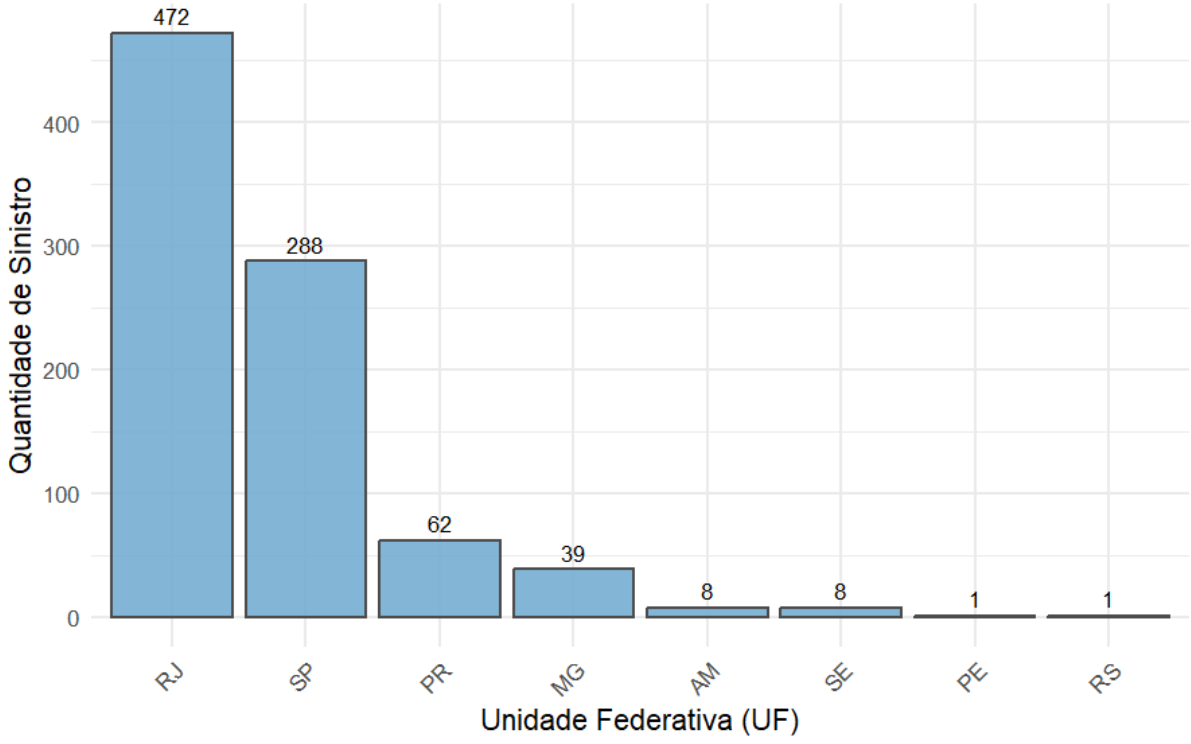
Fonte: A autora (2025)

Figura 14 – Distribuição de apólices por UF para o ramo Habitacional



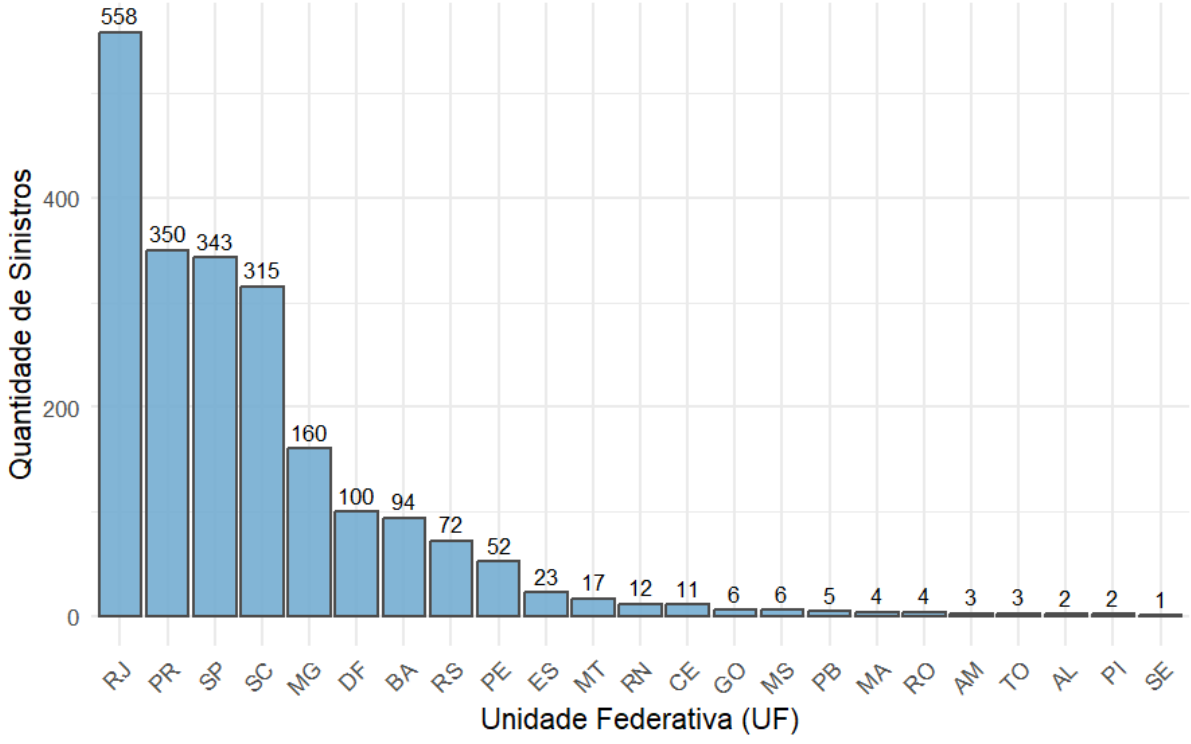
Fonte: A autora (2025)

Figura 15 - Distribuição de Sinistro por UF para o ramo Habitacional



Fonte: A autora (2025)

Figura 16- Distribuição de sinistros por UF no ramo de Responsabilidade civil



Fonte: A autora (2025)

4.2. Subscrição Adversa e Risco Moral

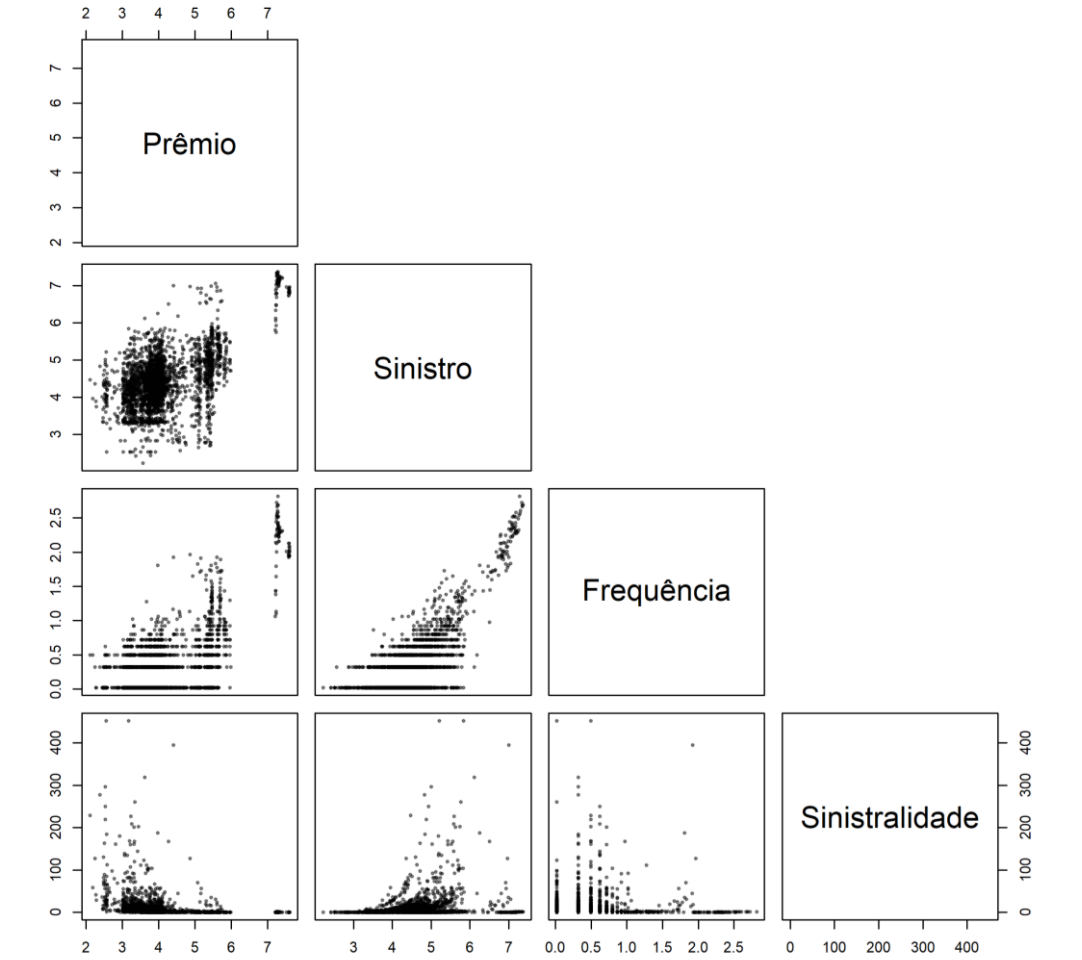
Nesta etapa, investigamos de forma mais aprofundada sobre o comportamento técnico e econômico da carteira, buscando identificar padrões de subscrição adversa e possíveis sinais de risco moral. A análise tem como objetivo compreender, por meio da clusterização, se há grupos de apólices com sinistralidade desproporcionalmente elevada em relação aos prêmios arrecadados, o que pode indicar falhas na precificação do risco ou assimetria de informações entre seguradora e segurado. A análise foca em métricas técnico-atuariais e utiliza características contratuais (UF e ramo) apenas para interpretação comparativa, portanto, apenas foram examinados dados que dispõem das informações completas.

As bases PREMIT, SINPAG E SINPEND foram consolidadas e agregadas por apólice-endosso, como uma chave, e para cada contrato foram obtidas as informações de prêmio total líquido, sinistro e frequência. Por sinistro consideramos o valor somado de todos os sinistros para uma mesma chave.

As variáveis UF e Ramo foram preservadas no *dataset* consolidado para uso apenas nas comparações posteriores; não foram usadas para formar os clusters. Durante o processo de análise, verificamos que a variável Importância Segurada apresentava inconsistências relevantes na base de dados e chegamos à conclusão que não poderia ser utilizada como medida válida de exposição de risco nem para o cálculo da taxa de prêmio, sendo necessário adotar outras métricas para comparação das análises sugeridas.

Antes da aplicação dos métodos de agrupamento, foi construída uma matriz de dispersão (Figura 17) entre os principais indicadores técnico-atuariais, considerando apenas observações com valores positivos de Prêmio, Sinistro, Sinistralidade e Frequência. O objetivo foi examinar visualmente as correlações e padrões de dependência entre as variáveis, identificar possíveis outliers e compreender a estrutura geométrica da base, servindo como etapa exploratória para a clusterização.

Figura 17 - Matriz de dispersão das variáveis técnicas



Fonte: A autora (2025)

Para melhorar a visualização e reduzir o efeito da grande amplitude numérica entre observações, as variáveis monetárias (Prêmio e Sinistro) foram convertidas para escala logarítmica ($\log_{10}$). Essa transformação preserva as relações relativas entre as variáveis, mas comprime os valores extremos, permitindo observar os padrões gerais sem que poucos contratos de alto valor distorçam o gráfico.

A análise inicial da matriz de dispersão mostra uma correlação positiva moderada entre prêmio e sinistro, sugerindo que contratos com prêmios mais altos tendem, em média, a apresentar sinistros maiores, embora exista bastante heterogeneidade entre os diferentes perfis de risco. Já a relação entre sinistro e frequência é a mais evidente, é possível observar que

conforme a frequência aumenta, o valor total dos sinistros cresce quase de forma linear, indicando a presença simultânea de riscos de frequência e severidade na carteira.

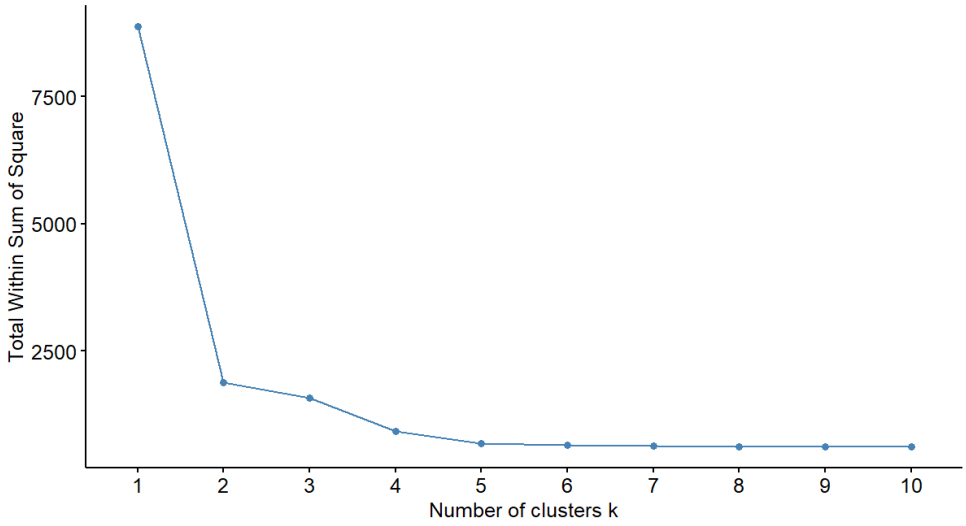
Por outro lado, as métricas de sinistralidade apresentam alta dispersão, especialmente em contratos com prêmios menores, em que pequenos sinistros podem gerar índices de sinistralidade extremamente elevadas. Esse comportamento reflete a existência de cauda longa e valores extremos, que impactam significativamente as médias.

De modo geral, a matriz revela uma carteira com ampla variabilidade de comportamento técnico, composta principalmente por contratos de pequeno porte e baixo risco, mas com subconjuntos concentrando perdas elevadas, justificando a aplicação do método de *K-means* para identificar grupos homogêneos de apólices com padrões distintos de desempenho e precificação.

As variáveis escolhidas para o agrupamento foram prêmio, sinistro e frequência, todas padronizadas por *z*-score (scale()), assegurando média 0 e desvio-padrão 1, para evitar dominância de variáveis monetárias (Prêmio e Sinistro) sobre contagens (Frequência).

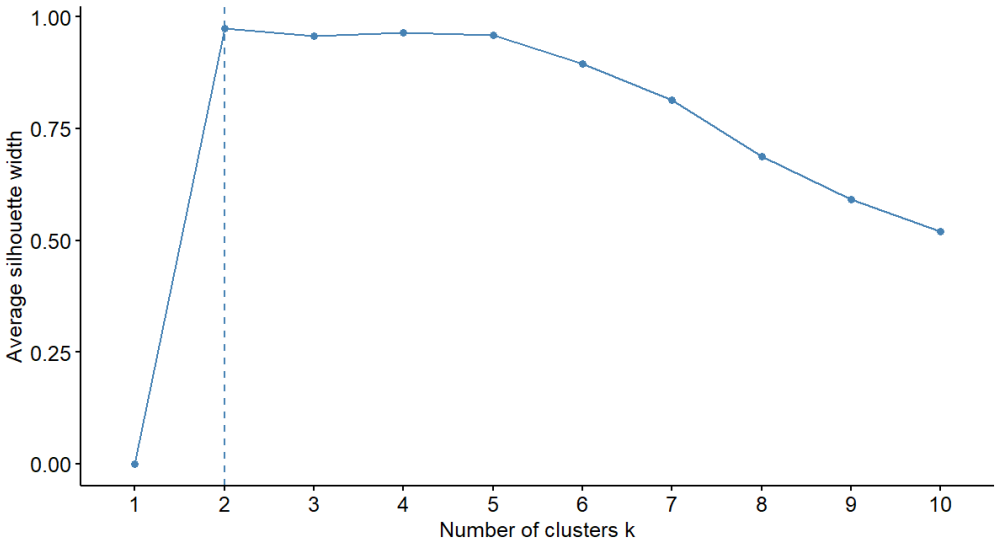
Para a determinação do número ótimo de clusters foram aplicadas a métrica de Cotovelo (WSS) e o Silhueta (ROUSSEEUW, 1987; JAIN, 2010). No gráfico do Cotovelo (Figura *18*), observa-se uma forte redução no WSS até k = 2, seguida por uma queda menor, porém ainda relevante, ao passar para k = 3. Apesar de o ponto de inflexão mais claro ocorrer em k = 2, a redução adicional obtida com três clusters indica que essa solução ainda agrega informação. Já no método da Silhueta (Figura *19*), o valor máximo ocorre em k = 2, com diminuição leve em k = 3. Assim, embora k = 2 seja a solução mais coesa segundo os critérios puramente estatísticos, a proximidade entre os valores permite adotar k = 3, favorecendo uma segmentação mais detalhada sem perda significativa de qualidade. Então, executou-se *k*-means com *k*=3.

Figura 18 – Gráfico do método de Cotovelo



Fonte: A autora (2025)

Figura 19- Gráfico do método Silhueta



Fonte: A autora (2025)

A segunda etapa da análise teve como objetivo identificar padrões de subscrição adversa e risco moral dentro da carteira, a partir da aplicação de técnicas de agrupamento baseadas nas variáveis técnico-atuariais. O procedimento adotado consistiu na execução do método K-

means, com três centróides (k = 3) e 25 inicializações (nstart = 25), utilizando as variáveis padronizadas Prêmio, Sinistro, Sinistralidade, Frequência e Severidade média.

Depois de identificar os clusters, foram calculadas medidas-resumo (como médias e medianas) para as variáveis técnico-atuariais de cada grupo. Para facilitar a interpretação, também foram criados gráficos, como boxplots de sinistralidade e um gráfico de barras com as médias por variável, que mostraram claramente as diferenças entre os perfis.

Como as variáveis monetárias apresentavam grande variação, aplicou-se a escala logarítmica nos gráficos de prêmio e sinistro. Além disso, foram feitas versões alternativas com ajustes nos limites dos eixos e painéis separados por cluster, permitindo uma visualização mais detalhada da dispersão dentro de cada grupo.

A análise de clusters revelou três perfis técnicos bem distintos dentro da carteira. O Cluster 1 representa um segmento de exposição intermediária: apólices com valores de prêmio e sinistro, além de frequência moderada, formando um grupo tecnicamente estável e proporcional ao volume de negócios. O Cluster 2 reúne os contratos de maior porte da carteira, caracterizados por prêmios, sinistros e frequências extraordinariamente elevados. São apólices caracterizadas por vigência longa e consequentemente maior exposição ao risco. Trata-se de um extrato de grande relevância financeira e que concentra parte considerável do risco assumido pela seguradora. Já o Cluster 3 compreende apólices de pequeno valor e baixa frequência, refletindo a parcela mais pulverizada da carteira, porém com maior variabilidade relativa em sua experiência de sinistros.

Comparando os Cluster 1 e 2 na Tabela *5*, visualiza-se que os valores dos sinistros do Cluster 1 são um pouco menos da metade dos valores dos sinistros totais do Cluster 2. Por outro lado, o Cluster 1 tem prêmios em proporções notadamente inferiores, a média e a mediana do valor do prêmio do cluster 1 é menos que um terço e aproximadamente 2,5% que os valores respectivos no Cluster 2. Em relação a frequência, observamos o mesmo comportamento, a média e a mediana representam aproximadamente 30% dos valores do Cluster 2. Tais fatos indicam serem apólices que, apesar de apresentar um menor número de sinistros (menor frequência) e menor prêmio, apresentam a severidade dos sinistros superior.

O Cluster 3, além de ser o cluster que representa 97,1% da carteira da seguradora, é o cluster que contém as apólices que apresentam em média 2,5 sinistros (mediana=2), com prêmios e sinistros totais claramente inferiores aos outros dois clusters. Em todos os clusters vemos a mesma assimetria positiva para o prêmio (média superior a mediana), enquanto em relação ao sinistro temos para o Cluster 1 e 2 mediana e médias quase iguais (leve uma

assimetria negativa e positiva, respectivamente) e para o Cluster 3, temos uma assimetria mais forte (a média é igual a 67,6 mil e a mediana 24 mil).

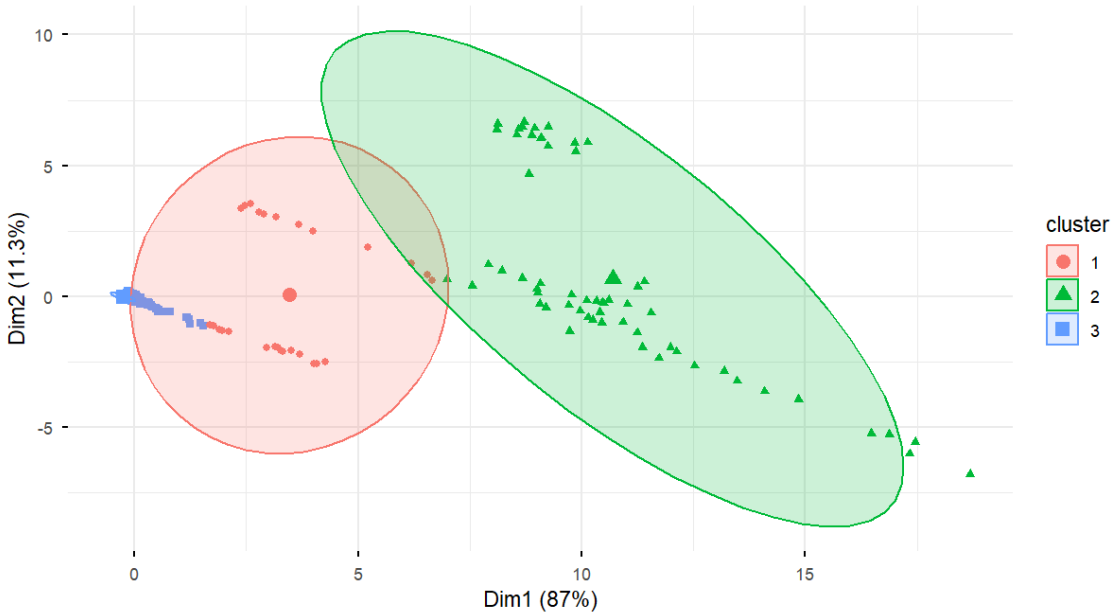
Tabela 5 - Medidas-resumo dos Clusters

		Prêmio		Sinistro		Frequência	
Cluster	**Nº de Apólices**	**Média**	**Mediana**	**Média**	**Mediana**	**Média**	**Mediana**
1	27	7.627.345,06	508.913,02	5.401.428,76	5.467.960,46	62,30	57
2	59	24.515.507,51	20.074.081,41	12.585.938,11	12.388.854,90	218,12	196
3	2873	61.017,18	8.243,72	67.636,49	24.441,36	2,53	2

Fonte: A autora (2025)

A estrutura desses grupos é bem ilustrada pelos gráficos produzidos, especialmente pelo mapa PCA, que sintetiza a maior parte da variabilidade dos dados nas duas primeiras componentes. Na Figura *20* observa-se que o Cluster 2 aparece visualmente como o grupo mais afastado dos demais, ocupando uma região associada a valores muito elevados de prêmio, sinistro e frequência. Essa posição reflete um conjunto de apólices de grande porte, com forte exposição ao risco e participação significativa no resultado técnico da seguradora. Já o Cluster 1 localiza-se numa faixa intermediária, representando apólices com volumes médios e comportamento técnico estável, trata-se de um grupo menos extremo, cuja dispersão moderada indica relativa homogeneidade interna. Por fim, o Cluster 3 concentra-se próximo à origem do gráfico, caracterizado por apólices de baixo valor, sinistros pequenos e baixa frequência, típicas de carteiras pulverizadas.

Figura 20 - Mapa PCA (Principal Component Analysis)

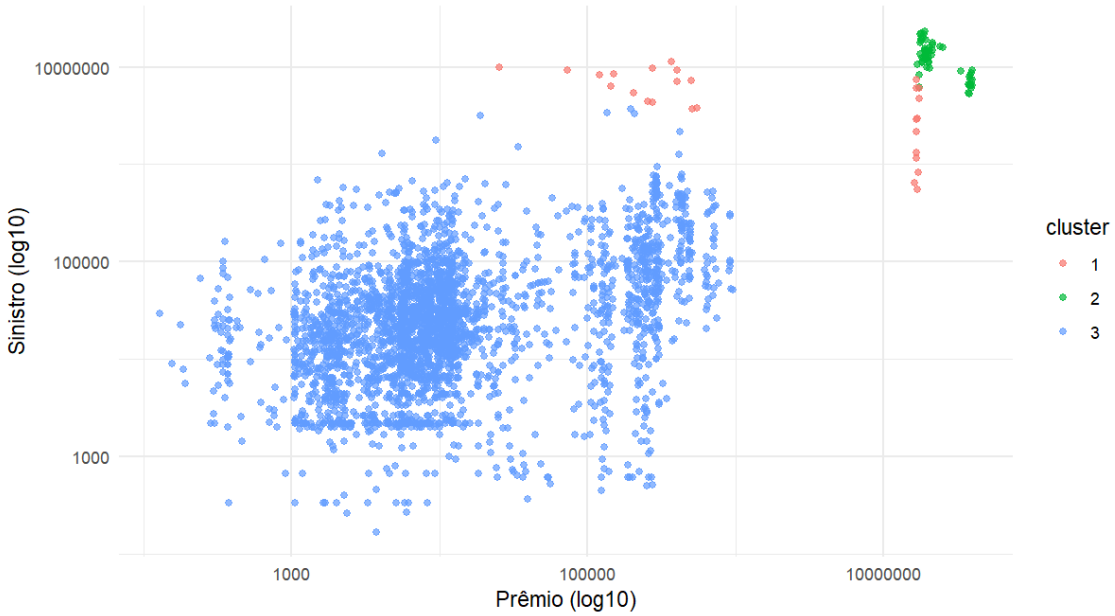


Fonte: A autora (2025)

O gráfico bivariado em escala log-log de Prêmio × Sinistro (Figura *21*) reforça essa estrutura: o Cluster 2 aparece claramente nos níveis mais elevados, enquanto o Cluster 3 se concentra nas menores magnitudes, e o Cluster 1 distribui-se de forma intermediária. A diferença entre os grupos torna-se ainda mais evidente quando se analisam as estatísticas descritivas. As médias e medianas do Cluster 2 confirmam sua natureza de alto risco e grande exposição; o Cluster 1 apresenta valores intermediários e distribuição mais equilibrada; e o Cluster 3 revela baixa materialidade econômica, com forte assimetria típica de ramos de pequeno valor.

Em conjunto, os gráficos e tabelas demonstram que os clusters representam diferentes perfis técnicos da carteira, permitindo identificar grupos de baixo, médio e alto risco. A segmentação revelou-se consistente, estatisticamente bem definida e alinhada com o comportamento observado nos dados, oferecendo uma base sólida para análises de subscrição, precificação e gestão de risco.

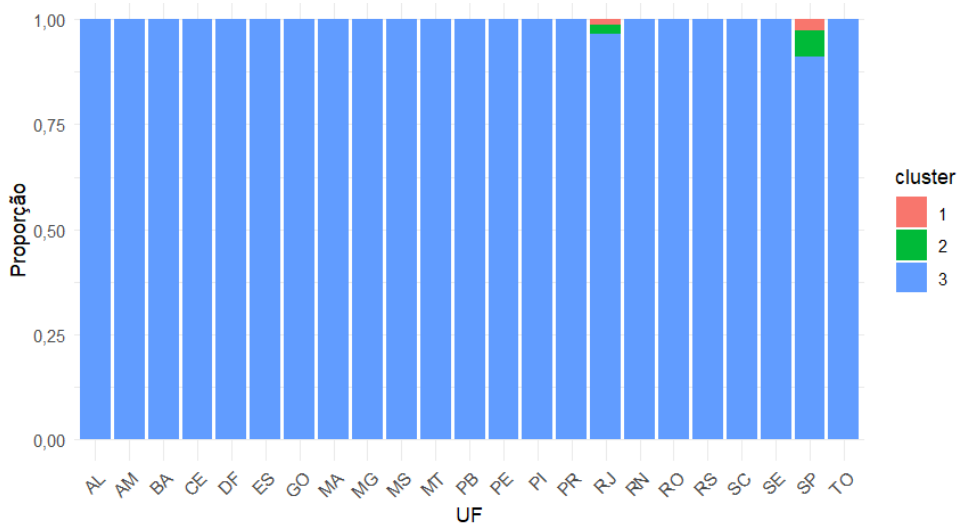
Figura 21– Distribuição Prêmio x Sinistro (log-log) por Cluster



Fonte: A autora (2025)

Após a formação dos clusters exclusivamente com variáveis técnicas, foram realizadas comparações com variáveis contratuais externas (UF de risco e ramo de seguro) que não participaram do processo de agrupamento, mas serviram para aprofundar a interpretação dos perfis identificados. A análise regional (Figura *22*) mostrou que o Cluster 3 domina praticamente todas as regiões do país enquanto os Clusters 1 e 2 surgem apenas em dois estados, sendo eles Rio de Janeiro (RJ) e São Paulo (SP). Essa configuração indica que a dinâmica técnica predominante da carteira é homogênea territorialmente, sem diferenciações significativas entre regiões no que se refere ao comportamento conjunto de prêmio, sinistro e frequência.

Figura 22- Distribuição dos Clusters por UF

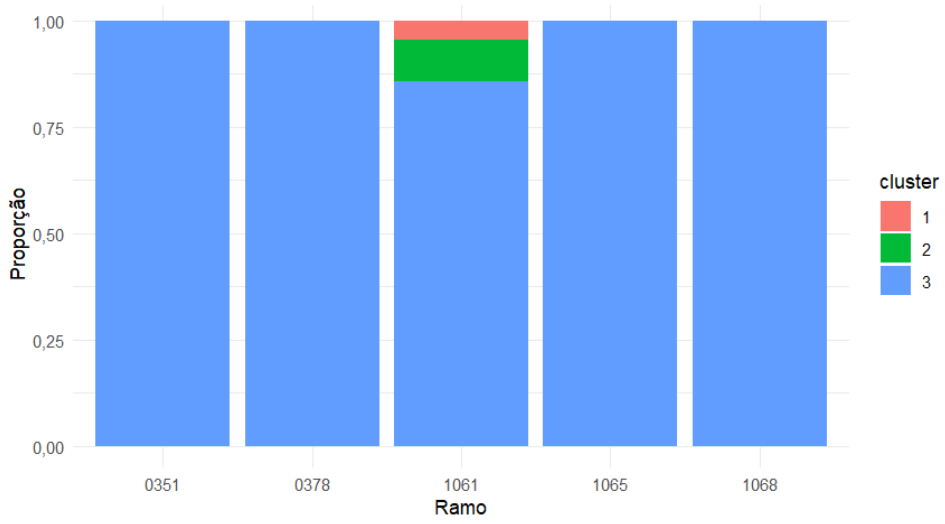


Fonte: A autora (2025)

No recorte por ramos de seguro (

Figura 23), verificou-se comportamento semelhante: o Cluster 3 permanece como classe majoritária em todos os ramos analisados, enquanto os Clusters 1 e 2 aparecem apenas dentro do ramo HAB MIP (1061), e ainda assim em proporções pequenas. Essa concentração estrutural revela que eventuais perfis diferenciados de risco não estão distribuídos de forma ampla entre os produtos, mas sim restritos a situações muito específicas.

Figura 23- Distribuição de Clusters por Ramo



Fonte: A autora (2025)

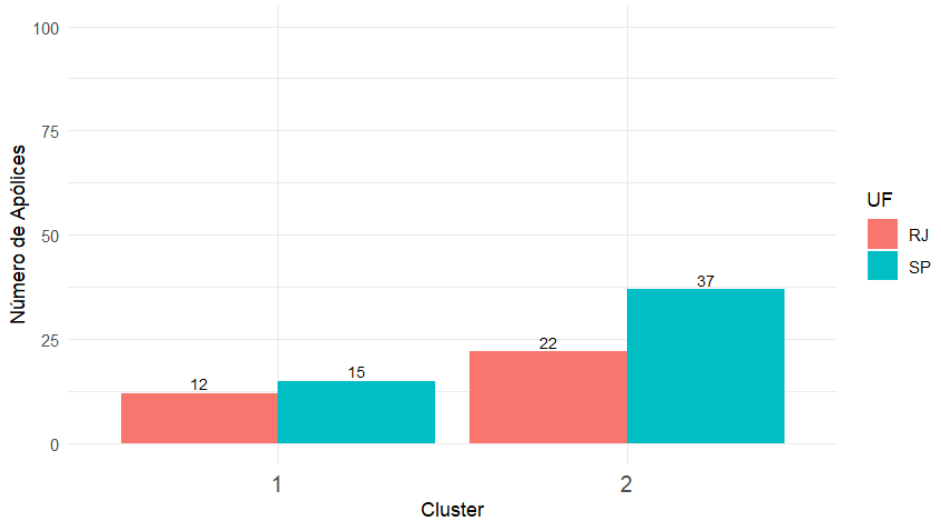
O cluster 1 e 2, como anteriormente mencionado são do ramo HAB MIP (1061), que é um seguro que cobre o saldo devedor do financiamento habitacional quando o segurado falece ou sofre invalidez permanente [ver Tabela *1*]. Isto é, tem a característica de que o segurado deve morrer ou estar permanentemente inválido para que possa registrar o sinistro, o que também não representa ser um ramo em que se apresente risco moral. Portanto, é um ramo que tem como características ser um seguro coletivo (logo uma frequência elevada não é sinal de risco moral) e é um seguro de severidade dos sinistros como o valor do financiamento, logo severidade elevada é prevista e a subscrição é feita de acordo.

Em conjunto, as figuras e tabelas demonstram que os clusters representam diferentes perfis técnicos da carteira, permitindo identificar diferentes grupos de risco. A segmentação revelou-se consistente, estatisticamente bem definida e alinhada com o comportamento observado nos dados.

A análise de clusters não evidenciou padrões consistentes de subscrição adversa ou risco moral. Isto é, os dois clusters minoritários apresentem exposição significativamente superior, sua ocorrência é concentrada e está associada a contratos de grande porte. A predominância das apólices no Cluster 3 indica homogeneidade técnica e ausência de deterioração intencional da frequência ou severidade de sinistros.

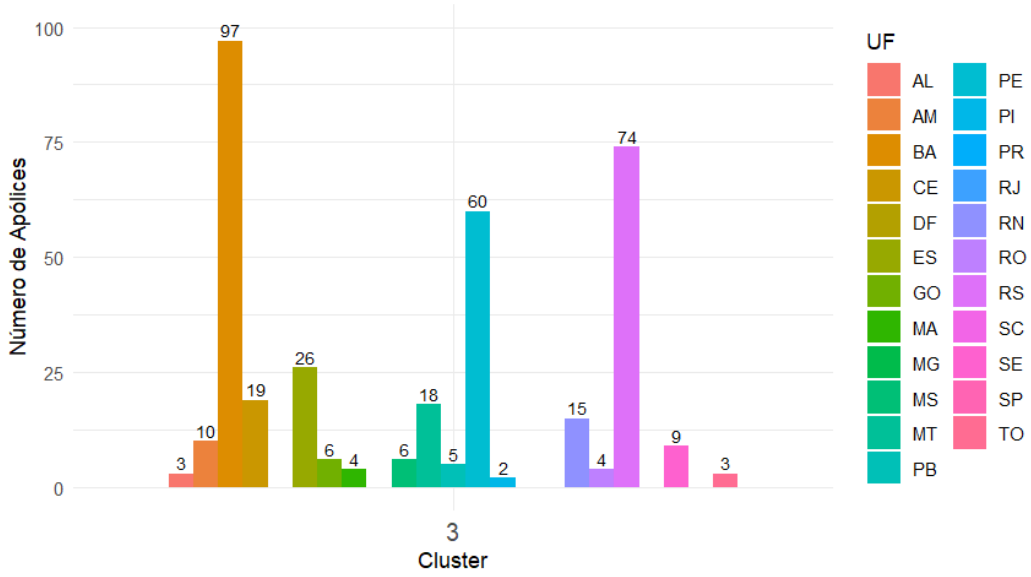
A Figura 24 apresenta as UFs dos Clusters 1 e 2, enquanto a Figura 25 para o cluster 3. A Figura 26 apresenta o boxplot de Prêmio, Sinistro e Frequência por Cluster e a Figura 27 o gráfico de perfil médio das variáveis por Clusters.

Figura 24 - UFs nos Clusters 1 e 2



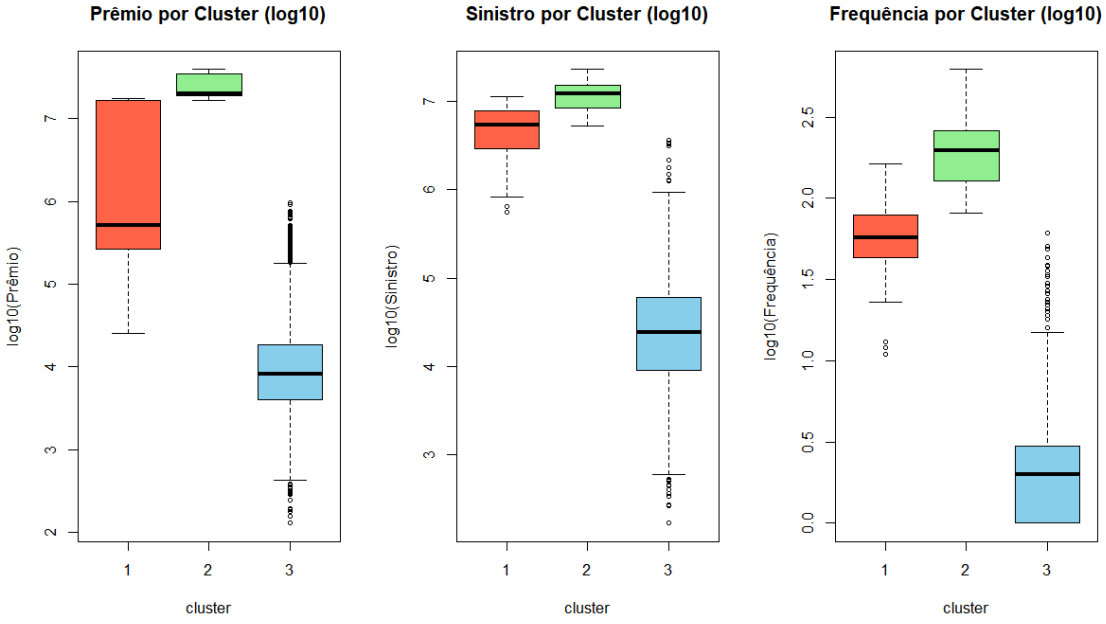
Fonte: A autora (2025)

Figura 25 - UFs do Cluster 3



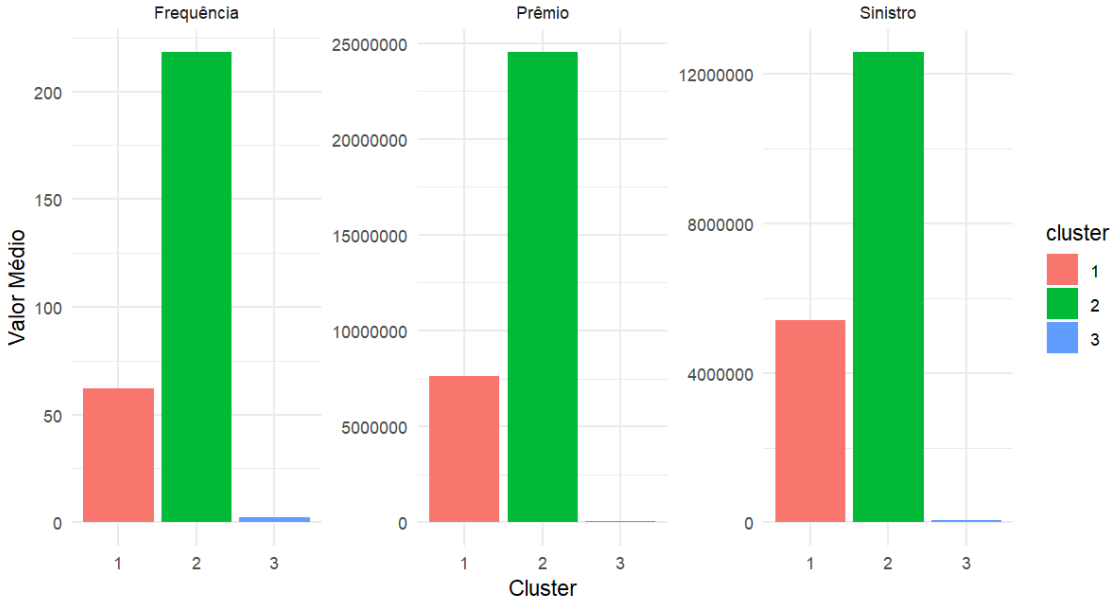
Fonte: A autora (2025)

Figura 26 - Boxplot de Prêmio, Sinistro e Frequência por Cluster



Fonte: A autora (2025)

Figura 27 - Gráfico de perfil médio das variáveis por Clusters



Fonte: A autora (2025)

4.3. Equidade na Precificação

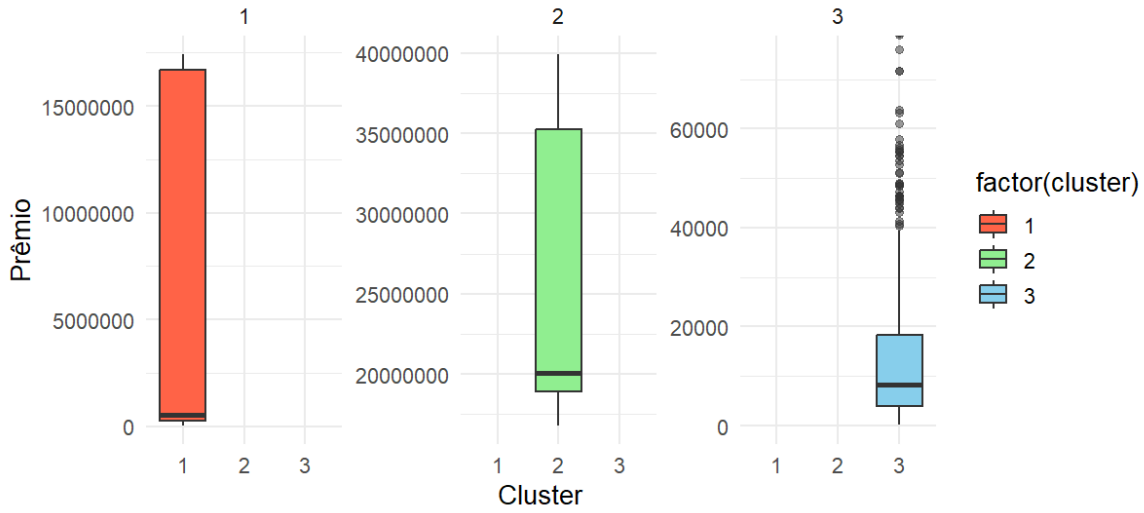
A terceira etapa da pesquisa teve como objetivo avaliar a equidade da precificação na carteira, isto é, verificar em que medida os prêmios cobrados refletem coerentemente o risco assumido pela seguradora. Para isso, foram consideradas exclusivamente as apólices-endosso já agregadas na etapa anterior, o que garante que cada observação represente uma unidade contratual completa, com seus respectivos valores consolidados de prêmio, sinistro e sinistralidade.

A partir desse conjunto, construiu-se uma variável de estratificação formada pela combinação entre ramo e UF de risco, de forma a definir perfis minimamente homogêneos para comparação. Essa etapa é essencial, pois tarifas distintas podem ser justificáveis entre ramos ou regiões diferentes, mas tornam-se indesejáveis quando ocorrem dentro de segmentos estruturalmente semelhantes.

Para conduzir essa avaliação, tomou-se como ponto de partida os clusters previamente identificados na análise 4.2. Foram elaborados *boxplots* comparativos para prêmios (Figura *28*) e Sinistros (Figura *29*), desta vez com escalas independentes para cada grupo, devido às diferenças expressivas de magnitude entre os clusters, evidenciando que os grupos realmente

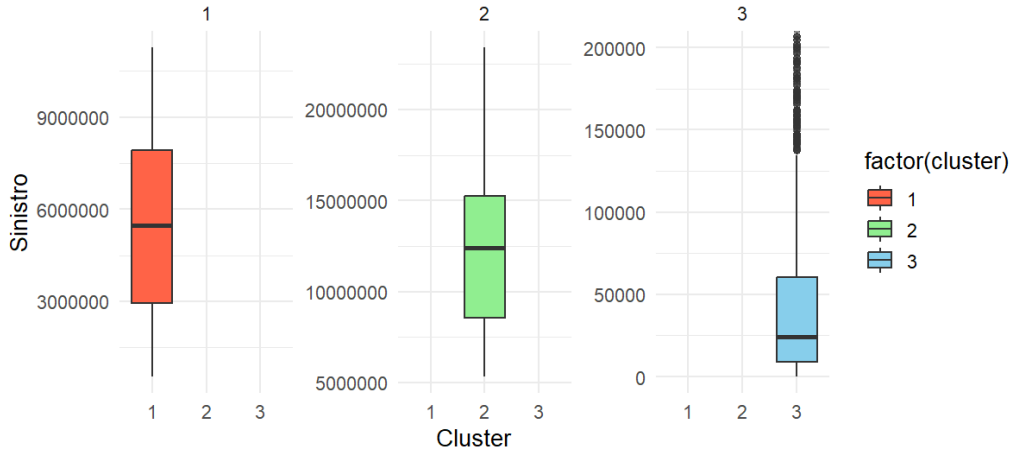
preservam estruturas distintas de prêmio e risco, um primeiro indicativo de que a segmentação anterior captura perfis tecnicamente diferenciáveis.

Figura 28- Boxplot de Prêmios por Clusters



Fonte: A autora (2025)

Figura 29 - Boxplot de Sinistros por Clusters



Fonte: A autora (2025)

Para confirmar essa diferença de maneira estatística, aplicou-se o teste não paramétrico de Kruskal–Wallis, adequado à natureza assimétrica das distribuições envolvidas. Esse teste foi escolhido em detrimento da ANOVA tradicional, uma vez que as variáveis envolvidas apresentaram distribuições assimétricas, com caudas longas e presença de outliers. O teste

avalia a igualdade das medianas entre três ou mais grupos, e sua estatística geral pode ser expressa como:

$$H = \left[\frac{12}{N(\mathrm{N}+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(\mathrm{N}+1)$$

Em que:

- N= número total de observações;
- k= número de grupos (clusters);
- n_i= tamanho do grupo i;
- R_i= soma dos postos atribuídos às observações do grupo i.

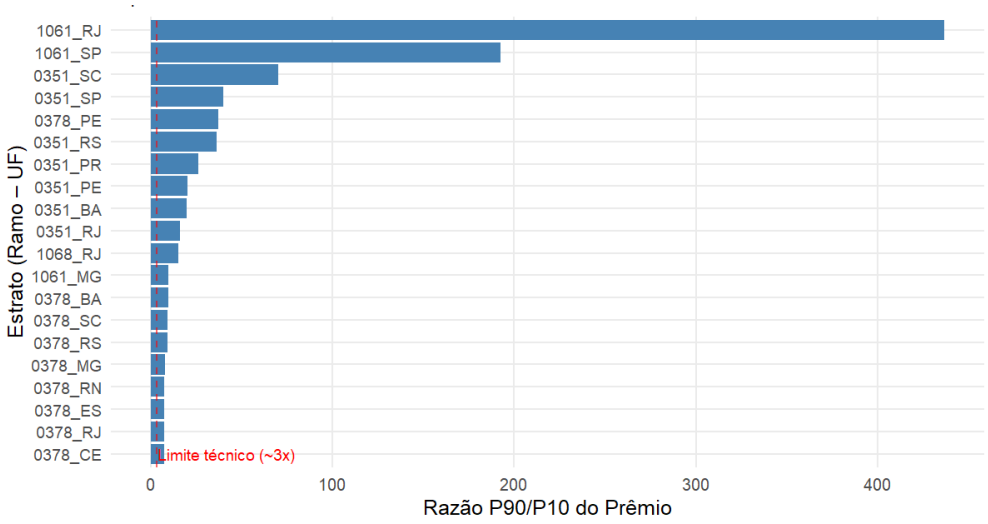
A hipótese nula é H_0: As distribuições do prêmio são iguais entre os clusters. Os resultados apontaram p-valor inferior a 10^{-51} e estatística de teste = 235,04, rejeitando a hipótese nula de igualdade entre as distribuições de prêmio dos clusters. Em complemento, o pós-teste de Dunn, ajustado pelo método de Bonferroni (DUNN, 1964), indicou que as diferenças são estatisticamente significativas nos pares que envolvem o Cluster 3, enquanto a comparação entre os Clusters 1 e 2 não apresentou diferença relevante. Isso demonstra que o Cluster 3 realmente se destaca dos demais, reforçando a validade da segmentação e a existência de perfis técnicos bem distintos.

Na sequência, a análise voltou para a dispersão dos preços dentro de cada estrato (ramo × UF). A intenção foi avaliar se apólices semelhantes do ponto de vista estrutural recebem preços compatíveis entre si. Para isso, foi calculada, em cada estrato, a razão entre o percentil 90 e o percentil 10 do prêmio, definida por:

$$\mathrm{Dispersão}_e = \frac{P90_e}{P10_e}$$

Essa razão indica quão distante estão os valores cobrados entre os segurados de maior e menor preço dentro do mesmo perfil. A variação observada mostra que em alguns casos, o prêmio do percentil 90 é mais de 100 vezes maior que o do percentil 10, chegando a ultrapassar 400 vezes em situações extremas. Diferenças tão grandes dificilmente se explicam por critérios técnicos. Em geral, indicam falta de padrão na precificação, subjetividade na subscrição ou até erros operacionais. Isso é preocupante porque distorções assim comprometem a equidade, o princípio de tratar riscos semelhantes de forma igual, e acabam gerando desequilíbrios dentro da carteira. A Figura 30 e a Tabela 6 apresentam a dispersão do Prêmio por perfil homogêneo.

Figura 30 – Dispersão do Prêmio: Top 20 estratos com maior razão P90/P10



Fonte: A autora (2025)

Tabela 6 – Dispersão do Prêmio por perfil homogêneo

Estrato	n	P90	P10	iqr_sinistro	Disperção
0351_AL	1	165.438.294	165.438.294	0.00000	1.000.000
0351_BA	18	121.045.166	6.148.475	4.518.036.992	19.687.022
0351_CE	2	20.693.449	13.994.644	6.813.108	1.478.669
0351_DF	1	545.679.859	545.679.859	0.00000	1.000.000
0351_ES	1	6.220.301	6.220.301	0.00000	1.000.000
0351_MG	6	169.778.311	54.813.792	1.539.023.238	3.097.365
0351_MT	1	60.493.052	60.493.052	0.00000	1.000.000
0351_PE	22	218.525.379	10.762.488	3.615.798.269	20.304.356
0351_PI	2	32.492.785	19.354.675	47.137.565	1.678.808
0351_PR	13	243.977.483	9.359.340	3.295.677.165	26.067.808
0351_RJ	19	66.439.493	4.198.244	4.199.592.812	15.825.544
0351_RS	18	211.050.644	5.844.244	2.160.556.274	36.112.565
0351_SC	20	300.225.086	4.281.404	1.226.878.939	70.123.047
0351_SP	11	280.043.789	7.046.265	4.137.583.927	39.743.579
0378_AL	2	3.851.368	3.851.368	482.595.154	1.000.000
0378_AM	3	108.951.492	13.946.290	901.734.890	7.812.220
0378_BA	79	134.351.819	14.318.715	2.327.143.967	9.382.952
0378_CE	17	21.788.128	3.043.147	3.097.626.334	7.159.737
0378_DF	100	137.427.111	29.408.758	4.845.085.748	4.672.999
0378_ES	25	129.445.066	17.644.635	2.082.239.450	7.336.228
0378_GO	6	54.586.532	12.100.259	356.763.738	4.511.187
0378_MA	4	167.266.687	120.915.426	8.240.854.963	1.383.336
0378_MG	155	123.992.826	16.087.076	2.751.822.484	7.707.605

0378_MS	6	112.044.213	41.395.633	2.233.463.731	2.706.667
0378_MT	17	117.715.937	16.679.528	3.745.726.369	7.057.510
0378_PB	5	104.614.168	55.136.246	1.560.916.662	1.897.376
0378_PE	37	132.988.926	3.595.710	2.237.638.430	36.985.447
0378_PR	355	121.217.174	17.933.941	4.003.535.922	6.759.093
0378_RJ	560	124.816.808	17.382.453	3.278.111.860	7.180.621
0378_RN	15	27.381.222	3.697.739	728.067.429	7.404.855
0378_RO	4	87.091.445	64.561.859	4.146.374.538	1.348.961
0378_RS	55	135.793.906	15.050.331	1.566.136.079	9.022.652
0378_SC	306	132.374.995	14.175.815	3.215.970.689	9.338.087
0378_SE	1	46.221.219	46.221.219	0.00000	1.000.000
0378_SP	336	139.317.414	31.398.586	3.372.328.283	4.437.060
0378_TO	3	58.554.030	54.694.604	634.554.182	1.070.563
1061_AM	7	1.864.184.633	1.830.227.280	8.614.458.653	1.018.554
1061_MG	23	2.808.367.002	293.316.638	30.810.494.407	9.574.523
1061_PE	1	361.980.076	361.980.076	0.00000	1.000.000
1061_PR	55	3.883.278.470	997.007.835	7.746.944.811	3.894.933
1061_RJ	267	167.681.758.867	383.806.110	26.509.128.184	436.891.843
1061_RS	1	132.089.759	132.089.759	0.00000	1.000.000
1061_SE	6	3.303.980.692	2.417.805.896	6.420.234.218	1.366.520
1061_SP	242	203.720.522.079	1.057.546.340	51.481.056.430	192.635.078
1065_MG	1	327.686.493	327.686.493	0.00000	1.000.000
1068_PR	1	4.816.975.311	4.816.975.311	0.00000	1.000.000
1068_RJ	125	4.979.865.453	328.686.417	11.687.841.990	15.150.810
1068_SE	2	370.362.735	367.302.431	538.678.054	1.008.332
1068_SP	2	1.119.322.068	686.087.798	13.358.986.871	1.631.456

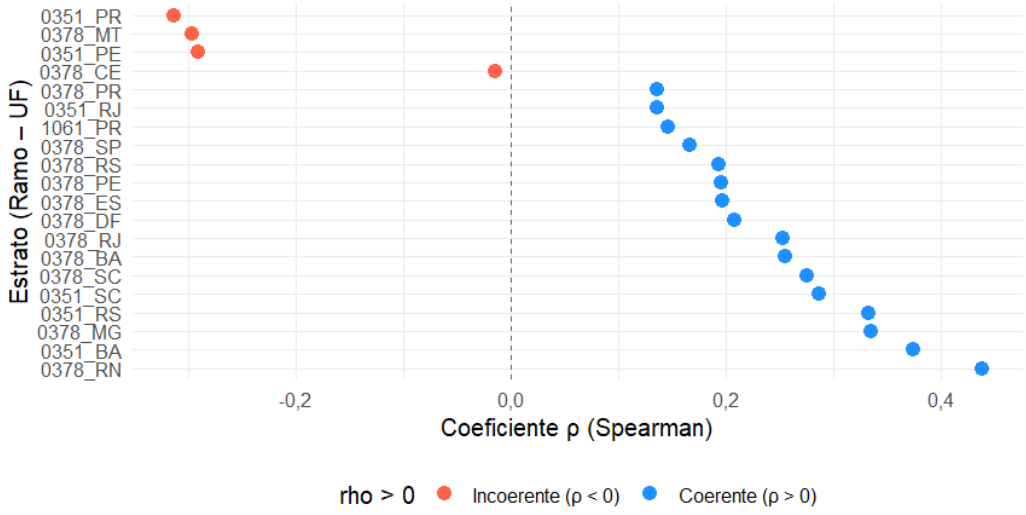
Fonte: A autora (2025)

Para aprofundar a análise, investigou-se a coerência entre preço e risco observados. Espera-se, do ponto de vista atuarial, que apólices com maior risco apresentem maiores sinistros, portanto, sejam tarifadas com prêmios mais elevados. Assim, dentro de cada estrato homogêneo, estimou-se a correlação de *Spearman* entre o prêmio total e os sinistros pagos. O coeficiente de *Spearman* é particularmente adequado nesse contexto porque não assume normalidade, sendo robusto à presença de outliers e de distribuições assimétricas, características comuns em dados de seguros. (KENDALL e GIBBONS, 1990)

De modo geral, o gráfico com os 20 casos com correlação extrema (Figura *31*) mostra que a maior parte dos estratos apresentou correlações positivas entre prêmio e sinistro, sugerindo que, na maior parte dos perfis homogêneos, apólices que pagam prêmios mais elevados também tendem a registrar sinistros mais altos. Esse comportamento é compatível

com uma estrutura tarifária minimamente coerente, na qual o preço acompanha a magnitude do risco observado. No entanto, a presença de alguns estratos com correlação negativa ou próxima de zero revela pontos de atenção importantes. Esses casos indicam que, em determinados grupos, prêmios maiores não correspondem a sinistros mais intensos, e sinistros altos podem estar ocorrendo justamente em apólices de prêmio reduzido. Esses desvios pontuais representam possíveis inconsistências na precificação ou na política de subscrição, sugerindo que a coerência preço–risco não é homogênea em toda a carteira.

Figura 31- Correlação entre Prêmio e Sinistro: Top 20 correlações mais extremas



Fonte: A autora (2025)

A Análise em 4.3 mostrou que, embora os clusters apresentem diferenças estatisticamente consistentes, indicando perfis de risco bem separados, ainda há indícios de desequilíbrios dentro dos estratos. A dispersão dos prêmios revelou variações muito superior ao esperado entre segurados com características semelhantes, sugerindo falta de uniformidade na tarifação. Já a correlação entre prêmio e sinistro apresentou, em geral, comportamento positivo, coerente com o esperado, mas com alguns casos isolados de correlação fraca ou negativa, indicando distorções pontuais. Em conjunto, os resultados indicam uma estrutura tarifária que funciona de forma adequada na média, mas ainda contém inconsistências que podem afetar a equidade e a eficiência técnica, reforçando a necessidade de ajustes pontuais nos critérios de precificação e na governança tarifária.

4.4. Probabilidade de ocorrência de sinistro em função do Prêmio

A quarta etapa do estudo tem como objetivo investigar se o valor do prêmio é um bom preditor da ocorrência de sinistros na carteira analisada, considerando também a influência das características contratuais, representadas pelo ramo do seguro e pela UF do risco. Essa análise complementa as investigações anteriores, que examinaram padrões de sinistralidade, dispersão de preços e coerência técnica de precificação, ao incorporar agora uma abordagem estatística voltada à modelagem da probabilidade de ocorrência do evento segurado.

A variável dependente definida nesta etapa é binária, assumindo valor 1 quando a apólice apresentou ao menos um sinistro no período de análise e 0 quando não houve qualquer ocorrência. Essa definição transforma o problema em uma tarefa de classificação binária, o que justifica o uso da regressão logística como método principal de modelagem. O modelo logístico permite estimar a probabilidade de ocorrência do sinistro a partir de um conjunto de variáveis explicativas e apresenta interpretação natural em termos de razão de chances (odds ratio), amplamente adotada em estudos atuariais e de risco.

A formulação geral do modelo segue a estrutura clássica, em que o logito da probabilidade de sinistro é modelado como combinação linear das covariáveis. Especificamente, utiliza-se como preditor principal o logaritmo natural do prêmio emitido, pois as distribuições de prêmio são altamente assimétricas e marcadas por valores extremos. A transformação logarítmica foi adotada como estratégia de modelagem para atenuar a assimetria da distribuição do prêmio e reduzir a influência de valores extremos, contribuindo para maior estabilidade numérica e melhor ajuste do modelo (HOSMER; LEMESHOW; STURDIVANT, 2013). Além disso, as variáveis categóricas referentes ao ramo do seguro e à UF de risco são incorporadas como fatores, permitindo captar diferenças estruturais de risco entre segmentações técnicas e geográficas. A interpretação é feita em termos de odds ratios, dados por:

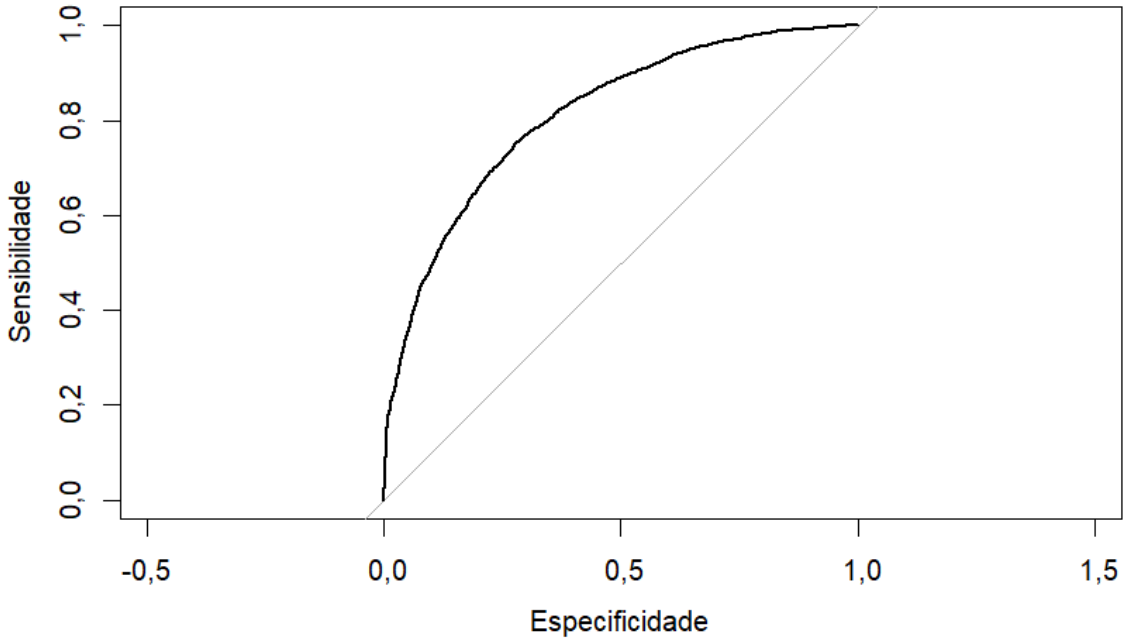
$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \log(\mathbf{1} + \text{Prêmio}_\text{i}) + \sum_k \gamma_k \times \mathbf{1}(\text{Ramo}_\text{i} = k) + \sum_j \delta_j \times \mathbf{1}(\text{UF}_i = j)$$

$$OR = e^{\beta}$$

Em que β coeficiente estimado associado a uma covariável no modelo.

Um odds ratio maior que 1 indica aumento da chance de sinistro, enquanto valores menores que 1 indicam redução. Os resultados da regressão logística indicam que o modelo apresenta desempenho preditivo satisfatório. A curva ROC estimada para o modelo (Figura *32*) evidencia um AUC próximo a 0,80, valor considerado elevado em estudos de risco não vida. (HOSMER; LEMESHOW; STURDIVANT, 2013). Isso significa que o modelo consegue separar com boa precisão as apólices que produziram sinistro daquelas que permaneceram sem ocorrência, revelando capacidade discriminatória consistente. O comportamento da curva sugere que a probabilidade prevista é sensível às diferenças reais presentes na carteira e que os preditores incluídos agregam informação efetiva sobre o risco. Podemos ver a Tabela 7 com o Summary da regressão logística.

Figura 32 - Curva ROC: Modelo Logístico



Fonte: A autora (2025)

Tabela 7 - Summary da regressão logística - probabilidade sinistro

term	estimate	std.error	statistic	p.value
(Intercept)	0,00	100,00	-0,26	0,80
log1p(Premio)	2,27	0,02	41,37	0,00
factor(COD_RAMO)0351	460304,01	100,00	0,13	0,90
factor(COD_RAMO)0378	849346,75	100,00	0,14	0,89
factor(COD_RAMO)1061	326261,45	100,00	0,13	0,90

factor(COD_RAMO)1065	113795,35	100,01	0,12	0,91
factor(COD_RAMO)1068	836474,46	100,00	0,14	0,89
factor(UF_RISCO)AM	1,92	0,66	0,98	0,33
factor(UF_RISCO)AP	0,00	362,58	-0,03	0,98
factor(UF_RISCO)BA	4,14	0,59	2,42	0,02
factor(UF_RISCO)CE	2,07	0,62	1,17	0,24
factor(UF_RISCO)DF	3,73	0,59	2,24	0,03
factor(UF_RISCO)ES	4,78	0,61	2,56	0,01
factor(UF_RISCO)GO	1,43	0,71	0,51	0,61
factor(UF_RISCO)MA	3,85	0,78	1,73	0,08
factor(UF_RISCO)MG	3,44	0,58	2,11	0,03
factor(UF_RISCO)MS	3,45	0,71	1,74	0,08
factor(UF_RISCO)MT	4,83	0,63	2,51	0,01
factor(UF_RISCO)PA	0,00	205,78	-0,06	0,95
factor(UF_RISCO)PB	1,84	0,73	0,83	0,41
factor(UF_RISCO)PE	2,51	0,59	1,55	0,12
factor(UF_RISCO)PI	4,84	0,92	1,71	0,09
factor(UF_RISCO)PR	4,51	0,58	2,59	0,01
factor(UF_RISCO)RJ	4,81	0,58	2,71	0,01
factor(UF_RISCO)RN	5,03	0,63	2,55	0,01
factor(UF_RISCO)RO	4,81	0,77	2,04	0,04
factor(UF_RISCO)RS	5,54	0,59	2,90	0,00
factor(UF_RISCO)SC	4,29	0,58	2,50	0,01
factor(UF_RISCO)SE	1,53	0,68	0,63	0,53
factor(UF_RISCO)SP	2,44	0,58	1,53	0,13
factor(UF_RISCO)TO	8,99	0,83	2,65	0,01

Fonte: A autora (2025)

Diferenças sistemáticas entre ramos e UFs também foram observadas. Ramos tradicionalmente associados a maior frequência de sinistros exibiram maior probabilidade ajustada, enquanto ramos mais estáveis apresentaram menor risco [ver Tabela 8]. Em relação às UFs, unidades com maior concentração populacional apresentaram probabilidades mais elevadas, coerentes com padrões de risco observados no mercado [ver Tabela 9].

Tabela 8 - Número de apólices e probabilidade observada por ramos

Ramo	n_apolices	prob_observada
1068	840	0,154762
1061	6052	0,099471
0378	160193	0,013053
1065	115	0,008696
0351	25712	0,00525
0313	1565	0

Fonte: A autora (2025)

Tabela 9 - Número de apólices e probabilidade observada por UF de Risco

UF_RISCO	n_apolices	prob_observada
TO	92	0,032609
MA	154	0,025974
RJ	41652	0,023312
RO	172	0,023256
SP	28329	0,020862
SE	445	0,020225
DF	5030	0,02008
MT	956	0,018828
PR	25062	0,016918
MS	382	0,015707
PI	134	0,014925
ES	1820	0,014286
MG	12965	0,014269
SC	24969	0,013056
RS	6086	0,012159
BA	8800	0,011023
AM	1395	0,007168
RN	2880	0,005208
GO	1208	0,004967
PE	13923	0,004309
PB	1607	0,003111
CE	12095	0,001571
AL	3887	0,000772
PA	320	0
AP	114	0

Fonte: A autora (2025)

A interpretação dos coeficientes estimados permite conclusões relevantes. Observa-se que valores mais altos de prêmio tendem a estar associados a maior probabilidade de ocorrência de sinistro, mesmo após o controle por ramo e UF. Como a importância segurada não pôde ser utilizada neste trabalho devido a inconsistências observadas na base original, o prêmio funciona como *proxy* da exposição ao risco. Assim, é plausível que maiores valores de prêmio reflitam apólices de maior complexidade, maior escopo de cobertura ou bens de maior valor econômico, o que se traduz em maior propensão à ocorrência de sinistros.

Os efeitos associados aos ramos e às UFs também apresentam um padrão coerente com as análises anteriores. Ramos com maior histórico de frequência ou volatilidade apresentam probabilidades médias mais elevadas, enquanto ramos tecnicamente mais estáveis exibem probabilidade menor. Do ponto de vista geográfico, observa-se que UFs mais popularizadas tendem a apresentar probabilidade média superior, possivelmente em razão de maior

concentração de bens segurados, maior densidade operacional e maior exposição a riscos associados ao ambiente urbano.

Para além da probabilidade de ocorrência de pelo menos um sinistro, estimada via regressão logística, a etapa seguinte da análise consiste em examinar a distribuição condicional da frequência de sinistros, isto é, como os sinistros se distribuem entre as apólices que tiveram pelo menos uma ocorrência. Esse procedimento é importante porque a regressão logística capta apenas a chance de ocorrência, mas não distingue apólices com sinistro único daquelas com múltiplos eventos.

A distribuição condicional permite estimar proporções como $P(N = 1 \mid N \geq 1)$, $P(N = 2 \mid N \geq 1)$, $P(N = 3 \mid N \geq 1)$e $P(N \geq 4 \mid N \geq 1)$. O comportamento observado revela que a maior parte das apólices sinistradas apresenta apenas um evento no período, enquanto ocorrências múltiplas são menos frequentes, embora variem conforme ramo e UF.

Tabela 10 - Frequência observada e prevista por ramo

COD_RAMO	n	freq_observada	freq_prevista
1061	602	29,43959979	18,69485397
1068	130	5,643871327	4,937483065
0351	135	2,133734821	2,551418504
0378	2091	1,926630375	2,055114064
1065	1	1,043674641	1,043674641

Fonte: A autora (2025)

Tabela 11 - Frequência observada e prevista por UF de Risco

UF_RISCO	n	freq_observada	freq_prevista
SP	591	13,99336354	10,66322777
RJ	971	11,93078117	7,103635305
MG	185	2,877157118	2,968827634
PR	424	2,156271192	2,813491016
MS	6	2,609186602	2,671173726
RO	4	2,609186602	2,654522461
RS	74	2,609186602	2,615008236
AM	10	2,087349282	2,550701768
AL	3	1,391566188	2,473893345
DF	101	2,128682931	2,173553346
SC	326	1,863247365	2,041180299
MT	18	1,855421584	1,956794049
PE	60	1,548117384	1,955913565
CE	19	1,592977083	1,932854256
BA	97	1,613929857	1,740242393

ES	26	1,445087964	1,596047782
MA	4	1,565511961	1,555560788
RN	15	1,321987878	1,486122457
GO	6	1,391566188	1,456647354
PB	5	1,252409569	1,246481617
SE	9	1,043674641	1,1986933
PI	2	1,043674641	1,053255307
TO	3	1,043674641	1,043817621

Fonte: A autora (2025)

Embora a abordagem condicional seja suficiente para descrever a distribuição empírica, uma modelagem mais estrutural da frequência poderia ser implementada por meio de modelos de contagem. Nessa classe de modelos, a distribuição de Poisson é o ponto de partida teórico, mas sua aplicação exige que a variância seja próxima da média (MCCULLAGH; NELDER, 1989). No caso analisado, observa-se superdispersão evidente na variável de frequência, com variância muito superior à média, o que viola os pressupostos do modelo de Poisson.

Em situações como essa, a distribuição Binomial Negativa é reconhecida como mais apropriada, pois incorpora um parâmetro de dispersão adicional que permite capturar essa variabilidade extra. Assim, embora o uso da Binomial Negativa não seja obrigatório para atingir os objetivos centrais desta análise, sua aplicação é tecnicamente recomendada e acrescenta rigor metodológico.

Os resultados obtidos permitem concluir que o prêmio, embora não seja uma variável puramente técnica de exposição como a importância segurada, ainda carrega informação relevante sobre o risco de ocorrência de sinistro. Isso ocorre porque em algumas carteiras, clientes de maior risco tendem a comprar coberturas mais amplas ou limites mais altos, o que pode elevar o prêmio, o qual reflete decisões de subscrição, segmentação geográfica e tarifação prévia, funcionando como um agregador das características de risco observadas pela seguradora.

Além disso, a modelagem confirma que fatores estruturais como ramo e UF exercem papel significativo na determinação do risco. A combinação entre regressão logística para ocorrência e análise condicional da frequência fornece um panorama completo do comportamento sinistros nas apólices estudadas e amplia a compreensão da carteira, contribuindo para diagnósticos de risco, avaliação de subscrição e discussões sobre coerência técnica na precificação.

5. CONCLUSÕES

Este trabalho teve como objetivo analisar uma carteira de seguros, com foco na avaliação da eficiência técnica, na equidade da precificação e na compreensão da heterogeneidade dos riscos da carteira. Para tanto, foi utilizada uma base de dados que integra informações de prêmios e sinistros no período de 2015 a 2025, permitindo a construção de indicadores como sinistralidade, frequência e severidade média, bem como a aplicação de métodos estatísticos e técnicas de segmentação.

A análise foi conduzida em quatro etapas principais. Na análise descritiva, foram observadas fortes assimetrias na distribuição dos prêmios e sinistros, além de elevada concentração em determinados ramos e unidades federativas. O ramo de RCP (0378) responde pela maior parte das apólices, mas baixa participação nos sinistros; o ramo HAB MIP (1061), ao contrário, concentra a maior parte dos sinistros com pequena participação no número de apólices. As distribuições de prêmios e sinistros exibem caudas longas, com muitos contratos de baixo valor e poucos contratos de alta materialidade técnica.

Na análise de eficiência, a sinistralidade, em geral, revela que a carteira há um bom desemprenho técnico. A verificação por ramo apresentou, para o Habitacional, uma predominância de sinistros próximo a zero, com poucos casos extremos, favorecendo uma maior estabilidade e equilíbrio nos resultados. Já no grupo de Responsabilidade civil, existe uma alta variabilidade técnica e maior risco agregado, já que as indenizações consomem maior parte dos prêmios arrecadados.

A segmentação por K-means com três grupos evidenciou perfis técnico-atuariais distintos. O Cluster 2 reúne apólices de grande porte com prêmios, sinistros e frequências muito elevados, representando alta exposição e impacto financeiro. O Cluster 1 ocupa posição intermediária, com volumes e frequências moderados e comportamento técnico mais estável. O Cluster 3 concentra a grande maioria das observações, com prêmios e sinistros baixos e frequência reduzida, refletindo carteira pulverizada. A leitura conjunta dos gráficos e das medidas-resumo não identifica um padrão claro de seleção adversa ou risco moral. Os dois clusters minoritários aparecem vinculados a contratos de maior porte, concentração prevista em linhas de negócio como o Habitacional coletivo, e a maior parte das apólices no Cluster 3 preserva homogeneidade técnica e resultados compatíveis com superávit para a sua maioria.

Dentro de perfis homogêneos por ramo e UF observou-se, em muitos casos, correlação de Spearman positiva entre prêmio e sinistro, o que é compatível com coerência técnica mínima. Ainda assim, a razão P90 sobre P10 do prêmio em diversos estratos alcançou valores muito

altos, sugerindo dispersão tarifária superior ao desejável. Há casos isolados com correlação baixa ou negativa, potenciais focos de revisão de critérios de subscrição e gestão tarifária.

O modelo logístico, com log do prêmio como preditor principal e fatores para ramo e UF, apresentou desempenho satisfatório, com AUC próxima de 0,82. O coeficiente positivo de log(prêmio) indica maior probabilidade de ocorrência de sinistro para apólices com prêmios mais elevados, mesmo após o controle por estrutura contratual e geográfica. Em termos atuariais, o prêmio funcionou como *proxy* de exposição em razão da indisponibilidade de uma medida de importância segurada confiável. Diferenças sistemáticas por ramos e UFs foram estatisticamente significativas em vários casos, alinhadas ao que se observou na análise descritiva. A avaliação condicional da frequência reforçou o padrão de baixa frequência com concentração no primeiro evento, e evidenciou superdispersão que recomenda, para avanços metodológicos, o uso de modelos de contagem mais flexíveis do que Poisson.

Os resultados sugerem três frentes de ação. Primeiro, manter monitoramento segmentado por clusters, dada a materialidade do Cluster 2 e o papel estabilizador do Cluster 3. Segundo, revisar estratos com dispersão tarifária extrema e correlação fraca entre preço e risco observado, priorizando ajustes de critérios de subscrição e padronização. Terceiro, explorar métricas adicionais de exposição na qual possível, para substituir o prêmio como *proxy* e refinar a avaliação de suficiência.

Uma das principais limitações do estudo foi a ausência de algumas variáveis relevantes na base de dados, como o valor da importância segurada e o detalhamento das coberturas contratadas e as datas de emissão e vigência; o que restringiu a análise de suficiência do prêmio frente à exposição real ao risco. Além disso, não foi possível realizar uma avaliação temporal detalhada para verificar se os casos de subprecificação ocorreram em períodos específicos ou se refletem práticas recentes da companhia.

Para trabalhos futuros, recomenda-se a incorporação de variáveis adicionais relacionadas à exposição ao risco e à estrutura contratual dos seguros, bem como o aprofundamento da análise evolutiva dos indicadores técnicos. Avaliações temporais poderão indicar se os padrões identificados decorrem de ajustes históricos da política de subscrição ou se ainda persistem. Além disso, recomenda-se analisar a dependência entre as variáveis com o uso de cópulas.

De forma geral, os resultados deste trabalho são relevantes para o aprimoramento da gestão de riscos da seguradora, especialmente no que tange à definição de critérios técnicos mais precisos para precificação e seleção de riscos, promovendo maior equilíbrio atuarial e sustentabilidade financeira da carteira.

REFERÊNCIAS

ALCOFORADO, Renata G.; EGÍDIO DOS REIS, Alfredo D.; POMMERET, Denys. Risk model with dependent frequency and severity for liability and housing insurance. *[Working Paper]*, 2025.

AMERICAN ACADEMY OF ACTUARIES. Loss Ratios and Health Coverages. Washington, D.C.: Loss Ratio Work Group, nov. 1998. Disponível em: https://www.actuary.org/sites/default/files/files/publications/lossratios%201198.pdf. Acesso em: 21 Nov. 2025.

CONGRESSIONAL RESEARCH SERVICE (CRS). Developing Scenarios for the Insurance Industry. Washington, D.C.: Library of Congress, 2022. Disponível em: https://www.jbs.cam.ac.uk/wp-content/uploads/2021/11/crs-developing-scenarios-for-the-insurance-industry.pdf. Acesso em: 21 Nov. 2025.

CORPORATE FINANCE INSTITUTE (CFI). Loss Ratio: Definition, Formula, Example, and Interpretation. Disponível em: https://corporatefinanceinstitute.com/resources/wealth-management/loss-ratio/. Acesso em: 19 out. 2025.

FERREIRA, Paulo Pereira. Modelos de precificação e ruína para seguros de curto prazo. Rio de Janeiro: Funenseg, 2002. 224 p. ISBN 8570523971. Disponível em: https://docvirt.com/docreader.net/DocReader.aspx?bib=bib_digital&pagfis=12690. Acesso em: 21 Nov. 2025

GRACE, Martin F. Loss Ratio Dynamics. Atlanta: Fox School of Business, Georgia State University, jan. 2021. Disponível em: https://www.fox.temple.edu/sites/fox/files/documents/Cummins%20Conference%202022/Loss-Ratio-Dynamics-01072022.pdf. Acesso em: 13 out. 2025.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression.* 3. ed. New York: Wiley, 2013. DOI:10.1002/9781118548387. Disponível em: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387. Acesso em: 21 Nov. 2025.

JAIN, A. K. Data Clustering: 50 Years Beyond K-means. Pattern Recognition Letters, v. 31, n. 8, p. 651-666, 2010. DOI: 10.1016/j.patrec.2009.09.011. Disponível em: https://www.sciencedirect.com/science/article/pii/S0167865509002323. Acesso em: 21 Nov. 2025.

KENDALL, M. G.; GIBBONS, J. D. Rank Correlation Methods. 5. ed. London: Edward Arnold, 1990. Disponível em: https://books.google.com.br/books?id=ly4nAQAAIAAJ. Acesso em: 21 Nov. 2025.

LANNA, Beatriz Duarte; SAES, Alexandre Macchione. Companhias de seguro na economia brasileira, 1889–1914. *Economia e Sociedade*, Campinas, 2020. DOI: 10.1590/1982-3533.2020v29n2art07. Disponível em: https://www.scielo.br/j/ecos/a/xxDbD885V6hZb5shf4FxYby/abstract/?lang=pt. Acesso em: 21 Nov. 2025.

LIMA, Ana Paula de Souza. *Avaliação da qualidade de subscrição de riscos das seguradoras brasileiras através do DEA (Data Envelopment Analysis).* 2008. 82 f. Dissertação (Mestrado em Economia) – Universidade Federal de Pernambuco, Recife, 2008. Disponível em: https://attena.ufpe.br/bitstream/123456789/3832/1/arquivo3421_1.pdf. 21 Nov. 2025.

LLOYD, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. (Original work published 1957). Disponível em: https://ieeexplore.ieee.org/abstract/document/1056489/. Acesso em: 21 Nov. 2025.

MAGALHÃES, Raphael de Almeida. *O mercado de seguros no Brasil.* Rio de Janeiro: Funenseg (Fundação Escola Nacional de Seguros), 1997. Disponível em: http://docvirt.com/docreader.net/WebIndex/WIPagina/Bib_Digital/9651. Acesso em: 21 Nov. 2025.

MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models.* 2. ed. Boca Raton: Chapman & Hall/CRC, 1989. DOI: https://doi.org/10.1201/9780203753736. Acesso em: 21 Nov. 2025.

NATIONAL BUREAU OF ECONOMIC RESEARCH (NBER). Moral Hazard and Adverse Selection in Health Insurance. *NBER Digest*, abr. 2016. Disponível em: https://www.nber.org/digest/apr16/moral-hazard-and-adverse-selection-health-insurance. Acesso em: 13 out. 2025.

NICHOLSON, W.; SNYDER, C. *Microeconomic Theory: Basic Principles and Extensions.* 10th ed. Mason: South-Western Cengage Learning, 2010. Disponível em: https://www.kufunda.net/publicdocs/nicholson_and_snyder_10th_ed.pdf. Acesso em: 21 Nov. 2025.

OECD. *Global Insurance Market Trends 2024.* Paris: OECD Publishing, 17 dez. 2024. DOI: https://doi.org/10.1787/5b740371-en. Acesso em: 5 jul. 2025.

PAULY, M. V. The Economics of Moral Hazard: Comment. *The American Economic Review*, v. 58, n. 3, p. 531–537, 1968. Disponível em: https://www.jstor.org/stable/1831861. Acesso em: 13 out. 2025.

POWELL, D.; GOLDMAN, D. Disentangling Moral Hazard and Adverse Selection in Private Health Insurance. *Journal of Econometrics*, v. 222, n. 1, p. 141–160, 2021. DOI: https://doi.org/10.1016/j.jeconom.2020.07.030. Acesso em: 21 Nov. 2025.

RIBEIRO, Paulo Gomes. *História do seguro: um resumo.* Rio de Janeiro: Fundação Escola Nacional de Seguros (atual Escola Nacional de Seguros), ago. 1994. Disponível em: https://bdlb.bn.gov.br/acervo/handle/20.500.12156.3/291560. Acesso em: 21 Nov. 2025.

ROCHA, Antônio Felipe Silvério da. *Modelagem GLM aplicada à atuária: uma utilização dos modelos lineares generalizados na precificação de seguros.* 2015. 64 f. Monografia (Bacharelado em Ciências Atuariais) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Fortaleza, 2015. Disponível em: http://repositorio.ufc.br/handle/riufc/31108. Acesso em: 21 Nov. 2025.

ROUSSEEUW, P. J. Silhouettes: *a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics*, 1987. Disponível em: https://doi.org/10.1016/0377-0427(87)90125-7. Acesso em: 21 Nov. 2025.

ROTHSCHILD, M.; STIGLITZ, J. E. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics*, v. 90, n. 4, p. 629–649, 1976. DOI: 10.2307/1885326. Disponível em: https://www.uh.edu/~bsorense/Rothschild&Stiglitz.pdf. Acesso em: 21 Nov. 2025.

SHI, Peng; ZHAO, Zhengjun. Regression for copula-linked compound distributions with applications in modeling aggregate insurance claims. *The Annals of Applied Statistics*, 2019. DOI: https://doi.org/10.48550/arXiv.1910.05676. Acesso em: 21 Nov. 2025.

SUPERINTENDÊNCIA DE SEGUROS PRIVADOS (SUSEP). *Relatório de gestão 2023: completo revisado ASCOM final.* Rio de Janeiro: SUSEP, 27 mar. 2024. Disponível em: https://www.gov.br/susep/pt-br/arquivos/arquivos-transparencia/relatorio_de_gestao_2023_completo_revisado_ascom_final.pdf. Acesso em: 3 jul. 2025.

WONGSUWATT, Sippavit et al. The influence of loss ratio on profitability of non-life insurance companies in Thailand: the moderating roles of company type. Journal of Community Development Research (Humanities and Social Sciences), v. 14, n. 1, p. 46–60, jan./abr. 2021. DOI: 10.14456/jcdr-hs.2021.5. Disponível em: https://www.journal.nu.ac.th/JCDR/article/view/Vol-14-No-1-2021-46-60. Acesso em: 21 Nov. 2025.

ZHANG, Y.; WALTON, N. Adaptive Pricing in Insurance: Generalized Linear Models and Gaussian Process Regression Approaches. *arXiv*, 2019. Acesso em: 6 ago. 2025. DOI: https://doi.org/10.48550/arXiv.1907.05381. Acesso em: 21 Nov. 2025.