

UNIVERSIDADE FEDERAL DE MINAS GERAIS – UFMG

INSTITUTO DE CIÊNCIAS EXATAS – ICEx

AMANDA SANTOS PANDOLFI

**Aplicação em dados de seguros de vida para previsão de sinistralidade:
Modelagem em séries temporais**

BELO HORIZONTE, MG – BRASIL

2022

AMANDA SANTOS PANDOLFI

**Aplicação em dados de seguros de vida para previsão de sinistralidade:
Modelagem em séries temporais**

Monografia apresentada ao curso de Ciências Atuariais, do Instituto de Ciências Exatas (ICEx), da Universidade Federal de Minas Gerais (UFMG), como requisito parcial para obtenção do título de Bacharel.

Orientadora: Jussiane Nader Gonçalves

BELO HORIZONTE, MG – BRASIL

2022

Agradecimentos

Primeiramente, quero agradecer à Deus pelo dom da vida e pela força em todos os momentos.

Dedico esse trabalho aos meus pais, Márcia e Paulo e a minha irmã, Letícia. Obrigada por serem a minha base e por me apoiarem durante todo o processo.

Agradeço em especial à professora Jussiane Gonçalves, minha orientadora, que sempre me inspirou, fazendo a estatística ser mais acessível e interessante para todos os alunos e despertando essa vontade de me aprofundar no mundo dos seguros, e sendo um suporte essencial durante toda a trajetória da minha monografia, da escolha do tema aos processos finais de escrita.

Agradeço também a todos os professores que tive a oportunidade de ter aula durante a graduação, obrigada por compartilharem tanto conhecimento comigo.

À toda minha família, que sempre foi sinônimo de amor e aconchego.

Aos meus amigos da faculdade, Felipe, Gabriela, Laura, Luiza, Thales e Tiago, que fizeram esses quatro anos serem mais divertidos e leves na medida do possível. Obrigada por dividirem comigo todos os momentos felizes e angustiantes dessa caminhada.

Às minhas amigas do coração, Lara, Ana Clara, Janine e Bruna, que são a família que eu escolhi e que sou grata por todo apoio e amor de sempre.

E, por fim, obrigada a todos que de alguma forma contribuíram para essa caminhada, tornando-a mais fluída e enriquecedora.

Resumo

A condição de não estaticidade dos seguros em geral se dá por se tratar de uma ciência que depende de muitos fatores sociais e econômicos para se sustentar. O Seguro de Vida Individual, em especial, é impactado por diversos agentes, desde políticas econômicas e culturais, até o desempenho das seguradoras quanto a oferta e demanda do seu produto e aspectos sociodemográficos que influenciam no comportamento do próprio consumidor do seguro de vida. Por isso, a análise de cada fator poderá facilitar a aceitação e a precificação do seguro. O conhecimento sobre o risco identificado é fundamental para que a avaliação seja realista e precisa. Logo, para garantir a solvência dessas empresas para que elas cumpram com os seus compromissos futuros, diversas premissas atuariais podem ser envolvidas no gerenciamento e precificação de seus riscos, como análise de frequência de sinistros, severidade, risco biométrico, análises de sensibilidade e subscrição, acompanhamento de sinistralidade, e cabe ao atuário responsável a adequar tais medidas a realidade da sua seguradora.

A proposta dessa dissertação objetiva desenvolver previsões para uma dessas premissas atuariais supracitadas, a sinistralidade, a partir de modelos de regressão e de Box-Jenkins. Para aplicação da metodologia considerou-se as oito maiores seguradoras em volume de prêmio disponíveis no Sistema de Estatísticas da SUSEP (SES) e os dados de 2011 a 2021 para montar a série temporal. O desempenho dos modelos propostos para o 1º semestre de 2022 foi comparado com os dados reais observados pelas seguradoras e disponibilizados pelo SES.

Palavras chave

Seguro de vida; sinistralidade; séries temporais; regressão; Box-Jenkins; previsão;

Abstract

The non-static condition of insurance in general is due to the fact that it is a science that depends on many social and economic factors to exist. Individual Life Insurance, in particular, is impacted by several agents, from economic and cultural policies, to the performance of insurers regarding the supply and demand of their product and sociodemographic aspects that influence the behavior of life insurance consumers themselves. Therefore, the analysis of each factor may facilitate the acceptance and pricing of insurance. Knowledge about the identified risk is essential for the assessment to be realistic and accurate. By that, to guarantee the solvency of these companies so that they can fulfill their future commitments, several actuarial assumptions can be involved in the management and pricing of their risks, such as analysis of frequency of claims, severity, biometric risk, sensitivity analysis and underwriting, loss ratio monitoring, and it is up to the responsible actuary to adapt such measures to the reality of his insurer.

The aim of this dissertation is to develop forecasts for one of the actuarial assumptions mentioned above, the loss ratio, based on regression and Box-Jenkins models. To apply the methodology, the eight largest insurers in terms of premium volume available in the Sistema de Estatísticas da SUSEP (SES) and data from 2011 to 2021 were considered to assemble the time series. The performance of the proposed models for the 1st semester of 2022 was compared with the actual data observed by the insurers and made available by the SES.

Keywords

Life Insurance; loss ratio; times series; regression; Box-Jenkins; forecast;

Lista de Figuras

Figura 1 – Gráfico da Sinistralidade, disponibilizado no SES	13
Figura 2 – Box-Plot dos Dados Originais	
Figura 3 – Box-Plot dos Dados Ajustados	20
Figura 4 – Gráfico da Série Sinistralidade entre 2011 a 2021	21
Figura 5 – Análise dos Resíduos M3	
Figura 6 – Histograma dos Resíduos M3	24
Figura 7 – ACF	25
Figura 8 – PACF	25
Figura 9 – Análise dos Resíduos ARIMA(1,0,1) com média diferente de zero.....	28
Figura 10 – Análise dos Resíduos ARIMA(1,0,1)(0,0,1)[12] com média diferente de zero	28
Figura 11 – Projeção da série pelo modelo ARIMA(1,0,1) com Intervalos de Confiança	29
Figura 12 – Projeção da série pelo modelo ARIMA(1,0,1)(0,0,1)[12] com Intervalos de Confiança.....	30

Lista de Tabelas

Tabela 1 – Resumo dos dados originais e ajustados.....	20
Tabela 2 – Comparação dos modelos de regressão	23
Tabela 3 – Real x Previsão por mês e intervalos de confiança de 95%	24
Tabela 4 – Comparação entre Modelos ARIMA e seus respectivos AIC	26
Tabela 5 – Real x Previsão por modelo e mês.....	29
Tabela 6 – Comparação entre os Erros Quadráticos Médios dos modelos de previsão	31

Sumário

1. Introdução	9
2. Revisão da Literatura	10
2.1 Contexto Histórico	10
2.2 Órgãos Reguladores	11
2.3 Especificidades do Seguro Vida Individual	11
2.4 Contexto Atual Brasileiro do Seguro de Vida Individual	12
2.5 Trabalhos Relacionados	13
3. Dados e Métodos.....	14
3.1 Dados.....	14
3.2 Previsão por Séries Temporais.....	14
3.3 Testes de Raiz Unitária	16
3.4 Métodos de Previsão por Regressão.....	16
3.5 Funções de Autocorrelação	17
3.6 Métodos de Previsão por Modelo Autorregressivo Integrado e de Médias Móveis – ARIMA	18
4. Resultados	19
4.1 Estatística Descritiva.....	19
4.2 Normalidade da Série.....	21
4.3 Testes de Raiz Unitária	22
4.4 Modelos de Regressão para Previsão por Séries Temporais	22
4.5 Modelo Autorregressivo Integrado e de Médias Móveis – ARIMA.....	25
4.5.1 ACF e PACF	25
4.5.2 Identificação do Modelo.....	26
4.5.3 Estimação do Modelo.....	26
4.5.4 Diagnóstico do Modelo	27
4.5.5 Previsões	28
4.6 Comparação entre os Modelos de Previsão propostos	30
5. Considerações Finais.....	31
6. Referências Bibliográficas	32

1. Introdução

O estudo sobre seguros não é estático por se tratar de uma ciência que depende de muitos fatores sociais e econômicos para se sustentar. Segundo a Escola Nacional de Seguros (ENS, antiga FUNENSEG), o Seguro de Vida Individual, em especial, é impactado por diversos agentes, tais como políticas econômicas e culturais em uma conjuntura macroeconômica, e o desempenho das seguradoras quanto a oferta e demanda do seu produto e aspectos sociodemográficos que influenciam no comportamento do próprio consumidor do seguro de vida, por exemplo, no cenário micro.

Conforme Huebner e Black (1976), existem certos fatores como: idade, sexo, constituição física, estado físico, histórico pessoal, herança genética, risco moral, prática de atividades de risco, profissão, ocupação, entre outros que devem ser analisados pela seguradora para aceitação do risco. A análise de cada fator poderá facilitar a aceitação e a precificação do seguro. O conhecimento sobre o risco identificado é fundamental para que a avaliação seja realista e precisa. Tais fundamentos são estritamente importantes pois estão ligados diretamente ao volume de prêmio necessário pela empresa para arcar com os possíveis montantes de sinistros futuros advindos dessas subscrições.

Os seguros de pessoas são produtos, em geral, de média ou longa duração e diversas premissas atuariais podem ser envolvidas no gerenciamento e precificação de seus riscos, como análise de frequência de sinistros, severidade, risco biométrico, análises de sensibilidade e subscrição, acompanhamento de sinistralidade, e cabe ao atuário responsável adequar tais medidas a realidade da sua seguradora.

O objetivo geral deste trabalho é prever as taxas de sinistralidade, uma premissa atuarial como supracitado, para o Seguro de Vida Individual por Cobertura de Morte - ramo 1391, definido pela Superintendência de Seguros Privados (SUSEP). Entende-se sinistralidade como a razão entre o Montante total de Sinistros Retidos pelo volume de Prêmio Ganho da empresa, dado intervalo de tempo. Isto é, em termos gerais, é a representatividade dos custos sobre a receita da empresa. Esse índice é fundamental pois a partir dele é possível analisar se a aceitação do risco e sua precificação estão sendo feitas de modo realista ou se precisam de novas metodologias para melhor adequação.

As séries temporais são pontos distribuídos em um período de tempo. De acordo com Chaves et al. (2014), a suposição que norteia a análise é que há um sistema causal mais ou menos constante, relacionado com o tempo, que exerceu influência sobre os dados no passado e pode continuar a fazê-lo no futuro. Este sistema costuma atuar criando padrões não aleatórios que podem ser detectados em análises estatísticas.

Segundo Morettin e Tolo (2006), os principais objetivos para analisar uma série temporal são:

- i) investigar o mecanismo gerador da série temporal;
- ii) fazer previsões dos valores futuros da série;
- iii) descrever o comportamento da série, tais como sazonalidade e tendência; e
- iv) procurar periodicidades relevantes nestes dados.

A aplicação mais comum ao analisá-las - e a ser utilizada nesse trabalho - é o prever o desempenho futuro da variável observada através de uma investigação do seu comportamento no passado. Dado Machado (2012), o melhor critério para escolher um modelo de previsão é a sua capacidade preditiva, ou seja, quão perto estão as previsões dos valores posteriormente observados.

Dado os expostos, os objetivos específicos deste trabalho são:

- I. Aplicar métodos estatísticos de previsão em dados mensais de seguradoras que atuam no ramo 1391, disponibilizados no SES (Sistema de Estatísticas da SUSEP), a fim de modelar as sinistralidades para o 1º semestre de 2022, considerando diferentes períodos de tempo.
- II. Analisar a aplicabilidade dos métodos de previsão da série a partir da comparação com os dados reais observados já disponíveis no sistema da SUSEP.

2. Revisão da Literatura

2.1 Contexto Histórico

A preocupação em resguardar e proteger a própria vida faz parte do instinto humano e está presente desde os primórdios da humanidade. Contudo, somente em meados da Revolução Industrial que foram estruturados os primeiros seguros de pessoas em moldes semelhantes aos que conhecemos hoje. Foram motivados pelas condições precárias de vida e trabalho, em que os trabalhadores da época eram obrigados a aceitar uma carga de trabalho excessivamente elevada, e era perigoso, pois não havia nada que os resguardassem. Além disso, os inválidos da época eram apenas demitidos e substituídos, isto é, esses casos ilustram perfeitamente a importância de um seguro contra invalidez. Ainda no século XIX, Otto Von Bismarck montou o primeiro modelo de Seguro Social, que abrangia seguros doença, invalidez e acidentes de trabalho, justamente em resposta às greves e pressões dos trabalhadores da época.

No Brasil, o primeiro seguro registrado foi o marítimo. Até então, as leis do “Código Comercial Brasileiro” (Lei nº 556, de 25 de junho de 1850), proibia o seguro de vida e permaneceu assim até 1855, quando a sua exploração foi autorizada. Segundo Lima (2018), dado grande interesse do mercado internacional no fluxo de seguros de vida no país, foi promulgada, em 5 de setembro de 1895, a Lei nº 294, dispondo exclusivamente sobre as companhias estrangeiras de seguros de vida, determinando que suas reservas técnicas fossem constituídas e tivessem seus recursos aplicados no Brasil, para fazer frente aos riscos aqui assumidos.

2.2 Órgãos Reguladores

A Superintendência de Seguros Privados, SUSEP, foi criada em 1966 e é responsável pela autorização, controle e fiscalização dos mercados de seguros, previdência complementar aberta, capitalização e resseguros no Brasil. Há também o órgão máximo, que estabelece as grandes diretrizes do setor de seguros, o Conselho Nacional de Seguros Privados (CNSP), cujo presidente é o ministro da Fazenda. O CNSP é também o órgão que decide as pendências no campo de seguros em última instância, regulamenta os seguros obrigatórios e estabelece os limites das operações de seguros no País.

2.3 Especificidades do Seguro Vida Individual

No Brasil, esse tipo de seguro pode ser dividido por cobertura de sobrevivência ou de risco - também chamado de cobertura por morte, e cada tipo possui suas próprias legislações específicas. A primeira, regida pela Resolução CNSP Nº 348/2017, é estruturada sob o regime financeiro de capitalização e é definida como a cobertura que garante o pagamento do capital segurado, pela sobrevivência do segurado ao período de diferimento contratado, ou pela compra, mediante pagamento único, de renda imediata. Possui diversos tipos de planos e modalidades. O presente trabalho analisará o Seguro de Pessoas por Cobertura de Risco. Este, é definido pela Resolução CNSP Nº 117/2004 como cobertura do seguro de pessoas cujo evento gerador não seja a sobrevivência do segurado a uma data pré-determinada. Em revés a sobrevivência, essa cobertura admite os regimes financeiros de Repartição Simples (RS), de Capitalização (RC) e de Capitais de Cobertura (RCC), e deve ser estruturada apenas na modalidade de benefício definido.

Para mais, apesar de existir as formas de contratação individual ou coletiva, este relatório explora apenas o Seguro de Pessoas Individual.

Ademais, como exemplo de seguros de pessoas com coberturas de risco, temos seguro de vida, seguro funeral, seguro de acidentes pessoais, seguro educacional, seguro viagem, seguro prestamista, seguro de diária por internação hospitalar e seguro perda de renda, seguro de diária de incapacidade temporária. Este trabalho analisará em especial o Vida, indicado pela SUSEP como Ramo 1391. A depender do plano escolhido, há um leque de coberturas oferecidas pelas seguradoras que podem ser vendidas juntas ou separadamente. Segundo a Circular SUSEP N° 395/2009, que estabelece a codificação dos ramos de seguro, as contabilizadas no Ramo de Pessoas Individual são Morte por Qualquer Causa (MQC), Invalidez Ocasional por Doença e Invalidez por Qualquer Causa (doença ou acidente), conforme o caso.

2.4 Contexto Atual Brasileiro do Seguro de Vida Individual

Segundo dados do SES, o Seguro de Pessoas segue crescendo expressivamente ao longo dos anos no Brasil e atualmente representa 16% das receitas acumuladas por segmento no país. Em especial, o Vida Individual apresentou aumento de aproximadamente 215% no volume de prêmios em 4,5 anos (jan/2017 a jun/2022), os quais estão na casa dos bilhões. Em relação aos sinistros, os valores absolutos são bem menores - estão nos milhões, como é o esperado, e o aumento não é proporcional aos prêmios. Nos mesmos 4,5 anos, esse crescimento foi de aproximadamente 58%. É a partir desses dados que se calcula o índice de sinistralidade, observando a razão entre os sinistros pelos prêmios. Essa taxa é essencial para as seguradoras porque por meio dela conseguem analisar se o grau de aceitação de risco e a precificação estão adequadas.

A sinistralidade, como pode ser observada na Figura 1 abaixo, não oscila muito, desconsiderando o aumento exorbitante durante a pandemia do Covid-19. Fora esse período, essa taxa variava entre 20% a 30%. No seu máximo, no ano de 2021, atingiu aproximadamente 44%, valor muito alto, mas condizente com o cenário.



Figura 1 – Gráfico da Sinistralidade, disponibilizado no SES

2.5 Trabalhos Relacionados

Em um estudo proposto por Vianna (2018), foi analisado a sinistralidade no contexto da saúde, aplicado a operadoras de planos privados de assistência à saúde médico-hospitalar, e objetivava a determinação dos fatores associados a esse índice e seus possíveis efeitos. Através desse trabalho, a autora concluiu que a sinistralidade se altera de acordo com a mudança de alguns fatores das operadoras, como a alteração do número de beneficiários, por exemplo. Além disso, apresentou medidas que podem ser tomadas visando a redução da sinistralidade, como a adoção de medidas preventivas com respeito à situação econômico-financeira, mantendo recursos próprios que auxiliam na capacidade da operadora em honrar com compromissos futuros e a importância da disponibilidade financeira, que auxilia a operadora a honrar com os compromissos a curto prazo.

Para mais, Correa (2022) apresentou os impactos causados nas seguradoras do ramo vida pelo aumento da sinistralidade devido a pandemia de Covid-19, ponto mencionado na subseção 2.4, Contexto Atual Brasileiro do Seguro de Vida Individual. Pelos resultados apresentados pela autora, as seguradoras conseguiram administrar e absorver esta alta dessa taxa de forma satisfatória e o resultado, no geral, foram índices operacionais positivos.

Ainda, Mori e Gonçalves (2016), aplicaram a metodologia de séries temporais em dados do sistema de saúde do Brasil e dos Estados Unidos, no período entre 2000-2012. Segundo as autoras, esse método de séries temporais foi aplicado apenas à série do número de pessoas que possuem plano de saúde privado no Brasil. Para os Estados Unidos aplicou-se uma regressão simples para a obtenção das previsões de novas adesões aos

planos de saúde. Ambas as séries apresentaram uma tendência positiva no número de beneficiários de planos de saúde.

3. Dados e Métodos

3.1 Dados

O mercado de seguros de vida individual não é bem distribuído, gerando uma falta de proporcionalidade no volume de prêmios entre as seguradoras que operam com esse tipo de serviço. Dentre as 43 entidades que operam com esse seguro disponíveis no SES, as duas maiores concentram aproximadamente 68,24% de todo o montante de prêmios. Logo, para esse trabalho, foram selecionadas empresas com receita de prêmios maior ou igual a 0,5 bilhão. Por esse filtro, passaram 8 empresas, que tiveram seus nomes omitidos para preservar suas imagens. Tais informações estão disponíveis no próprio painel de inteligência do mercado de seguros da SUSEP.

Além disso, as variáveis selecionadas para compor o banco foram Competência, Sinistros Retido e Ocorrido, Prêmio Ganho e Sinistralidade. Nas previsões em si, só foram utilizadas a Competência e a Sinistralidade, porém, como essa última é calculada a partir do Prêmio Ganho e do Sinistro, elas também foram incluídas na base. São definidas de acordo com a SUSEP como:

- Prêmio Ganho: Prêmio Emitido + Variação das Provisões Técnicas.
- Montante de Sinistros: Sinistros Avisados + Despesas + Parcela de Cosseguro Aceito - Parcela de Cosseguro Cedido - Salvados e Ressarcimentos + Serviços de Assistência + Variação das Provisões Técnicas.

Ademais, como o estudo se trata de dados distribuídos em um período de tempo, foram considerados todos os meses de competências disponíveis no SES, isto é, a partir de janeiro de 2011. Portanto, foram incluídas 132 observações mensais ou 11 anos completos, de 2011 a 2021. Além disso, para fins de comparação, como supracitado nos objetivos específicos, foram recolhidas as informações do 1º semestre de 2022 também.

3.2 Previsão por Séries Temporais

Uma série temporal é o conjunto de observações $\{Y(t), t \in T\}$ em que Y é a variável de interesse e T é o conjunto de índices. Pode ser dividida em:

- Discreta: $T = \{t_1, \dots, t_n\}$ e é denotado por Y_t .

- Contínua: $T = \{t_1 < t < t_2\}$ e é denotado por $Y(t)$.
- Multivariada: $Y_t = (Y_{1t}, \dots, Y_{kt})'$ em que $t \in T$.

Segundo Shumway et al. (2011), a correlação óbvia introduzida pela amostragem de pontos adjacentes no tempo pode restringir severamente a aplicabilidade de muitos métodos estatísticos convencionais tradicionalmente dependentes da suposição de que essas observações são independentes e identicamente distribuídas. A abordagem sistemática a qual responde às questões matemáticas e estatísticas colocadas por essas correlações de tempo são comumente referidas como análise de séries temporais.

Uma previsão é hipótese ou conjectura sobre o futuro. Dado Antunes et al. (2015), a previsão de valores futuros ou passados pode ser feita por regressão linear, considerando-se o melhor ajuste dos pontos da série. A previsão também pode utilizar procedimentos estatísticos mais complexos, como o método ARIMA (*autoregressive integrated moving average*). O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série.

As séries temporais podem ser compostas dos seguintes fatores:

- tendência, que é o comportamento de longo prazo da série;
 - sazonalidade, que são flutuações periódicas nos valores da variável;
 - ciclos, são aquelas que provocam oscilações de subida e de queda nas séries, de forma suave e repetitiva, ao longo da componente de tendência;
 - erro ou variação irregular, que ocorre devido a efeito de um dos outros fatores.
- Nas componentes de erro aparecem flutuações de período curto e com deslocamento inexplicável.

Os modelos utilizados para descrever séries temporais são processos estocásticos, controlados por leis probabilísticas. Denota-se como T um conjunto arbitrário, um processo estocástico é uma família $\{Y_t, t \in T\}$ tal que para cada $t \in T$, Y_t é uma variável aleatória. Um processo estocástico está determinado quando são conhecidas suas funções de distribuição de probabilidade conjuntas, porém, como estas não são conhecidas e dispõe-se de apenas uma amostra do processo (a série temporal observada) assumem-se os pressupostos de estacionariedade e ergodicidade do processo estocástico.

Segundo Morettin e Tolo (2006), esses fatores quanto à variabilidade das observações podem ser definidos como:

- Estacionariedade, quando as suas estatísticas não são afetadas por variações no tempo. Uma série será estacionária se dois primeiros momentos forem constantes ao longo do tempo, ou seja $E[Y_t] = \mu$ e $\text{Var}[Y_t] = \sigma^2$.
- Ergodicidade, se apenas uma realização do processo estocástico é suficiente para se obter todas as suas estatísticas. É uma propriedade referente à relação entre a média temporal de um processo estocástico calculada a partir de uma realização temporal. Um processo é dito ergódico em média se suas médias temporais e amostrais coincidem quando T tende a infinito. Todo o processo ergódico também é estacionário, pois uma realização de um processo não estacionário não poderá conter todas as informações necessárias para a especificação do processo.

3.3 Testes de Raiz Unitária

Neste teste, a presença de uma ou mais raízes unitárias indica não estacionariedade no comportamento de uma série histórica, ou seja, os valores tendem a aumentar com o transcorrer do tempo, assumindo diferentes padrões.

Diniz (1998) afirma que uma série temporal é estacionária se os dados aleatórios oscilam em torno de um valor constante. Isto se constata quando os parâmetros de distribuição de probabilidade: média $E(Y_t) = \mu$, que mede o valor médio dos dados; a variância $\text{var}(Y_t) = \sigma^2$, que mede o grau de dispersão dos dados em relação ao valor médio; e a covariância γ_t , que mensura o grau de dispersão entre um dado e seu subsequente, são fixos e constantes ao longo do tempo (Gujarati, 2006).

Segundo Granger e Newbold (1974), ao se estimar uma regressão, mesmo que os valores dos respectivos testes t de student sejam significativos e que o coeficiente de determinação da regressão (R^2) seja elevado, ainda assim, corre-se o risco de se obter uma regressão espúria quando uma variável apresenta raiz unitária. Isso porque os pressupostos estatísticos de que a média e a variância devem ser constantes ao longo do tempo são violados, comprometendo, dessa forma, os resultados obtidos com a utilização dos modelos de regressão. Uma série que não seja estacionária, a previsão para o futuro não será eficiente, pois as variações de valores ao redor da média se afastam significativamente.

3.4 Métodos de Previsão por Regressão

Em um processo de regressão linear simples, supõe-se que $f(\cdot)$ qualquer pode ser aproximada por uma reta, tal qual

$$f(x) = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

em que β_0 é o intercepto, β_1 é a inclinação da reta e $\varepsilon_i \sim N(0, \sigma^2)$. O objetivo do modelo é ser uma aproximação simplificada da relação real entre as variáveis de interesse.

A regressão linear no contexto de séries temporais utiliza o método de mínimos quadrados para projetar uma equação linear que melhor se encaixe nos dados apresentados e assim fazer a previsão dos próximos meses. Assume Y_t como uma saída ou séries temporais dependentes, para $t = 1, \dots, n$. Essa variável está sendo influenciada por uma coleção de entradas possíveis ou séries independentes, tais como $X_{t1}, X_{t2}, \dots, X_{tq}$, consideradas como entradas fixas e conhecidas. Logo, essa relação pode ser expressa por

$$Y_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_q X_{tq} + \varepsilon_t$$

em que $\beta_1, \beta_2, \dots, \beta_q$ são coeficientes de regressão fixos desconhecidos e ε_t é um erro aleatório ou processo de ruído e $\varepsilon_i \sim N(0, \sigma^2)$.

3.5 Funções de Autocorrelação

A autocovariância é a covariância entre Y_t e o seu valor Y_{t-k} separado por k unidades de tempo e denota-se por:

$$\gamma_k = Cov[Y_t, Y_{t-k}] = E([Y_t - \mu][Y_{t-k} - \mu])$$

para $k = 0, \pm 1, \pm 2, \dots$

Autocorrelação, ou *Autocorrelation Function* (ACF), é a autocovariância padronizada. Serve para medir o comprimento e a memória de um processo, ou seja, a extensão para a qual o valor tomado no tempo t depende daquele tomado no tempo $t-k$ ou $t+k$ e denota-se por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov[Y_t, Y_{t-k}]}{\sqrt{Var(Y_t)Var(Y_{t-k})}}$$

Essa ideia de autocorrelação pode ser estendida, isto é, pode-se medir a correlação entre duas observações, eliminando a dependência dos termos intermediários. A autocorrelação parcial, *Partial Autocorrelation Function* (PACF) é denotada por:

$$Cov[Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-(k+1)}]$$

3.6 Métodos de Previsão por Modelo Autorregressivo Integrado e de Médias Móveis – ARIMA

De acordo com Shumway et al. (2011), uma regressão clássica, como proposta no tópico 3.2, muitas vezes não é suficiente para explicar todas as dinâmicas interessantes de uma série temporal. Em vez disso, a introdução da correlação como um fenômeno que pode ser gerada por meio de relações lineares defasadas leva a propor os modelos autorregressivo (AR) e autorregressivo de média móvel (ARMA). Esses, apesar de complexos, ajustam-se aos fatores sazonais e de tendência, estimam parâmetros adequados de peso, testam o modelo e repetem o ciclo quando adequado.

O modelo ARIMA, *autoregressive integrated moving average*, proposto pelos estatísticos George Box e Gwilym Jenkins, é uma difusão do modelo geral ARMA, aplicado a casos em que os dados são não estacionários. Essa questão da não estacionaridade pode ser resolvida a partir de diferenciações, o que corresponde a parte integrada (I) do modelo. Ambos modelos, ARIMA e ARMA, podem ser usados para prever o futuro de uma série, dado seus pontos passados. Para esses espaços de tempo, Box-Jenkins indicam um número de no mínimo 40 a 50 pontos. Os dados utilizados neste trabalho atendem esse requisito dado que o tamanho da amostra é 132, como supracitado. A previsão de Box-Jenkins é de maior utilidade quando se acredita que os fatores subjacentes que causam a demanda por produtos, serviços, receita e, neste caso, a taxa de sinistralidade se comportem no futuro da mesma maneira que no passado.

A parte Auto Regressiva (AR) dos modelos é estruturada por Y_t , que representa os valores atuais da série, que podem, na teoria, ser explicados em função de p , que significa os valores passados. Logo, é expresso por AR(p) e denota-se por:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

em que Y_t é estacionária e os coeficientes $\varphi_1, \varphi_2, \dots, \varphi_p$ são constantes (logo $\varphi_p \neq 0$). Assume-se que $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d. Portanto, essa parte do modelo indica que a variável é regressada em seus próprios valores defasados.

Do outro lado, assumindo os ruídos brancos et estrutura-se a parte de Médias Móveis, com ordem q . Dessa forma, é expresso por MA(q) e denota-se por:

$$Y_t = \varepsilon_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

em que θ_1, θ_2 até θ_q são parâmetros e $\theta_q \neq 0$. Novamente, assume-se que $\varepsilon_i \sim N(0, \sigma^2)$, salva indicação contrária. Logo, indica que o erro de regressão é na verdade uma combinação linear dos termos de erro, cujos valores ocorreram contemporaneamente e

em vários momentos no passado. Vale ressaltar que, diferente do AR(p), a parte MA(q) é estacionária para quaisquer valores dos parâmetros θ .

Ademais, a parte integrada (I) indica que os valores de dados foram substituídos com a diferença entre seus valores e os valores anteriores e este processo diferenciador pode ter sido realizado mais de uma vez, como explicado acima. Seguindo neste trabalho, na parte de resultados, será mostrado que não é este o caso, isto é, não foi preciso diferenciar a série pois ela já é estacionária.

Finalmente, as partes se juntam, formando os modelos ARMA e ARIMA, a depender da estacionaridade. Ressalta-se que quando $q = 0$, o modelo é chamado apenas de Auto Regressivo de ordem p , AR(p), e quando $p = 0$, é Médias Móveis de ordem q , MA(q). Portanto,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

em que os parâmetros p e q são chamados ordens autorregressivas e de médias móveis, respectivamente. Se Y_t tiver uma média μ diferente de zero, define-se

$$\alpha = \mu(1 - \Phi_1 - \dots - \Phi_p)$$

logo, o modelo poderá ser reescrito como

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

em que $\varepsilon_i \sim N(0, \sigma^2)$ independentes.

Neste trabalho, para calcular a previsão para os valores futuros a partir do modelo descrito, foi utilizada a função `arima`, do pacote `stats`, no software R. Essa aplicação ajusta um modelo ARIMA a uma série temporal univariada.

4. Resultados

4.1 Estatísticas Descritivas

Os dados originais de Sinistralidade possuíam dois outliers negativos e dois valores maiores do que 1, sendo, inclusive consideravelmente maiores que o terceiro quartil. Essas taxas correspondem aos meses de jan/11, dez/11, mai/16 e dez/17. Dessa forma, a fim de gerar a melhor previsão possível com os dados disponíveis, esses outliers foram ajustados por médias aritméticas. Para os outliers maiores do que 1, foi utilizado a média dos 10 maiores valores de sinistralidade menores do que 1. Em revés, para os outliers negativos, foi utilizado a média dos 10 menores valores de sinistralidade maiores

do que zero. O resumo dos novos dados em comparação aos originais pode ser observado na Tabela 1 abaixo.

Tabela 1 – Resumo dos dados originais e ajustados

	Sinistro		Sinistralidade	
	Originais	Ajustados	Originais	Ajustados
Mínimo	- 4.046.445,00	2.335.441,00	-0,0800	0,0800
1° Quartil	9.465.237,00	9.465.237,00	0,2100	0,2100
Mediana	15.674.779,00	15.674.779,00	0,2600	0,2600
Média	20.821.577,00	20.900.212,00	0,2940	0,2852
3° Quartil	25.194.276,00	25.194.276,00	0,3225	0,3225
Máximo	95.432.888,00	95.432.888,00	1,4900	0,6400
Desvio Padrão	18.232.735,00	18.134.014,00	0,1750	0,1197

Observa-se que os valores mínimo e máximo de Sinistralidade ficaram mais próximos da mediana e média sem outliers. Vale ressaltar, contudo, que mesmo com o ajuste proposto para os valores acima de 1, há ainda algumas taxas altas que se distanciam da realidade dos demais dados, como pode ser analisado com a relação entre o terceiro quartil, que representa até 75% dos dados até 0,32 e o máximo igual 0,64. A efetividade desses ajustes de outliers pelas médias pode ser melhor visualizada através dos boxplots da sinistralidade dos dados com e sem outliers, respectivamente, abaixo.

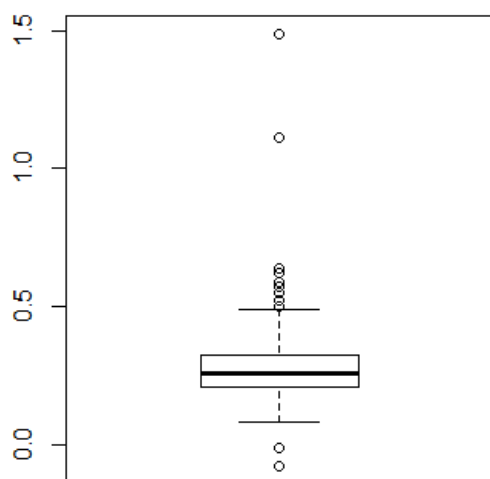


Figura 2 – Box-Plot dos Dados Originais

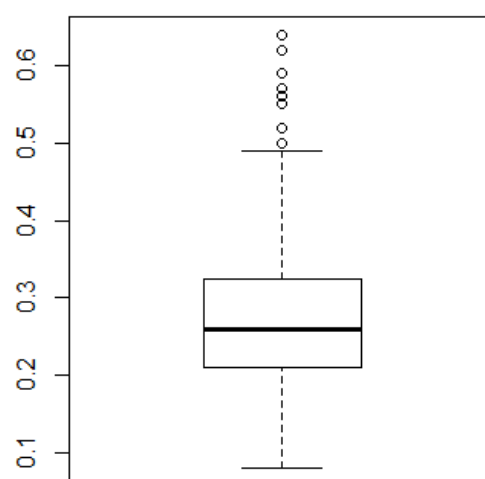


Figura 3 – Box-Plot dos Dados Ajustados

Como predito, pode-se observar através das Figuras 1 e 2 que o máximo dos *boxplots* é aproximadamente 0,5. Logo, os valores acima disso também são tratados como outliers mas serão considerados na previsão por serem mais frequentes.

Ademais, através dos dados ajustados, pode-se concluir que as taxas de sinistralidade observadas são, em maioria, menores do que 0,35, o que é um ponto muito positivo para as seguradoras estudadas. Isto é, a proporção entre sinistro ocorrido, o quanto a empresa deve pagar, e prêmio ganho, receita que entrou, é consideravelmente baixa, o que impacta diretamente na aptidão da empresa e retenção de lucros, e auxilia na comprovação da qualidade da subscrição e eficácia dos cálculos de precificação dos riscos.

A partir desse ponto, para as demais análises, foi utilizada a base com os dados ajustados. Abaixo, pode-se observar a série completa pelo tempo.

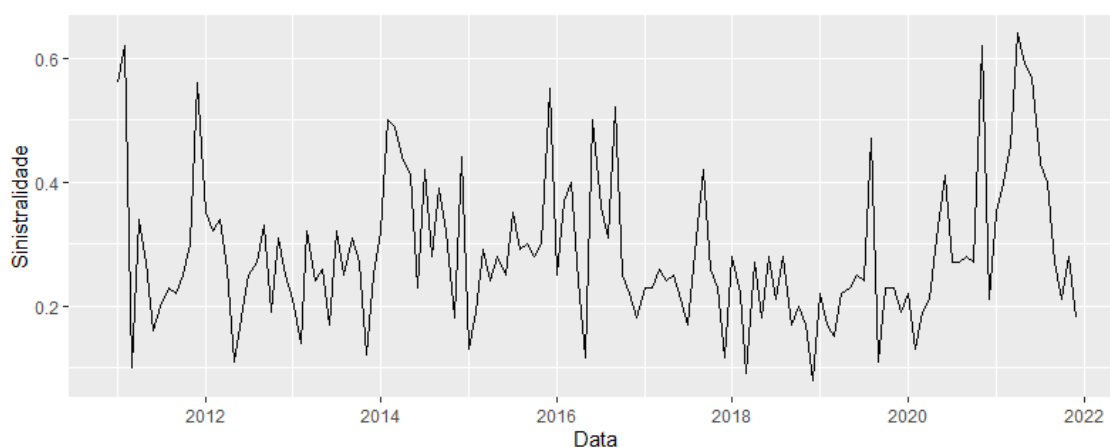


Figura 4 – Gráfico da Série Sinistralidade entre 2011 a 2021

Pelo gráfico acima, é possível notar que o fator da estacionaridade, essencial para o cálculo das previsões por séries temporais, provavelmente é atendido, dado que a série não aparenta apresentar nenhuma variação constante. A frente, essa hipótese será comprovada a partir dos testes de raiz unitária.

4.2 Normalidade da Série

Para testar a normalidade dos dados de Sinistralidade, a fim de averiguar se a distribuição de probabilidade associada à série pode ser aproximada pela distribuição normal, foi utilizado o teste de Shapiro-Wilk. A partir do p-valor do teste e considerando

um nível de significância de 5%, concluiu-se que a hipótese nula dos dados serem normais é rejeitada.

Diante disso, os dados da série foram transformados a partir do logaritmo e o teste Shapiro-Wilk foi aplicado a essa nova variável. Essa transformação tende a suprimir flutuações maiores que ocorrem em partes da série em que os valores subjacentes são maiores. Novamente considerando um nível de significância de 5%, a partir da série transformada, há evidências para se aceitar a hipótese nula de normalidade dos dados.

4.3 Testes de Raiz Unitária

Para essa análise, o primeiro teste adotado foi o ADF (*Augmented Dickey Fuller*, ou Teste de Dickey Fuller aumentado). A hipótese nula desse teste é que a série tenha raiz unitária, ou seja, não é estacionária. Se o valor da estatística, calculada intrinsecamente no teste ADF, for maior que o valor absoluto tabulado por Dickey-Fuller, a hipótese nula é aceita e, portanto, a série é não estacionária. Foi utilizada a função `adf.test` do pacote `tseries` no software R. Dado que $\tau\pi < \tau_3$ para os dados testados, é possível rejeitar a hipótese de que $\pi = 0$, logo não possui raiz unitária e é estacionária. Ou seja, conclui-se que a variável é $I(0)$. O número de defasagens foi determinado pelo AIC (*Akaike Information Criterion*, ou Critério de Informação de Akaike).

Ainda, proposta por Nusair (2003), o cruzamento dos testes ADF e o Kwiatkowski Philips Schmidt & Shin (KPSS) assegura uma conclusão mais precisa sobre a estacionariedade da série. Em revés, a hipótese nula do KPSS é que a série temporal é estacionária. Portanto, a fim de diminuir a incerteza do teste ADF realizado, foi utilizado também o teste KPSS, a partir da função `PP.test` também do pacote `tseries` no software R. O p-valor foi de 0,06076, portanto, considerando um nível de significância de 5%, não há evidências suficientes para rejeitar a hipótese nula de estacionariedade.

4.4 Modelos de Regressão para Previsão por Séries Temporais

Nessa seção, foram testados três modelos diferentes. No primeiro, a série Sinistralidade foi ajustada pelo período de tempo t (correspondendo a jan/11 a dez/21) mais o quadrado de t , correspondendo a tendência. O segundo modelo adiciona ao modelo 1 o fator de sazonalidade. E, por fim, o terceiro adiciona ao modelo 2 o termo autoregressivo de ordem 1 aos termos de erro et . Segue abaixo uma tabela comparativa,

levando em consideração a normalidade dos resíduos pelo teste Shapiro-Wilk, a estatística Durbin-Watson e o AIC.

Tabela 2 – Comparação dos modelos de regressão

	AIC	Durbin-Watson	P-Valor DW	P-Valor SW
Modelo 1	147,90	1,5144	0,004	0,0394
Modelo 2	165,47	1,5232	0,004	0,1330
Modelo 3	160,47	2,0535	0,862	0,1246

Primeiramente, a estatística Durbin Watson (DW) é um teste para autocorrelação de primeira ordem dos resíduos da regressão e sempre terá um valor variando entre 0 e 4. Um valor de 2 indica que não há autocorrelação detectada na amostra. Valores de 0 a menos de 2 indicam autocorrelação positiva e valores de 2 a 4 significam autocorrelação negativa. Em primeira instância, ao observar os valores dos três modelos, conclui-se que o Modelo 3 é o mais adequado, por ter valor DW igual a 2,05 enquanto os demais são menores do que 2. Porém, para confirmar, o p-valor também será levado em consideração. Portanto, usando um nível de 5% de significância, apenas há evidências de aceitar a hipótese nula do modelo 3, como esperado.

Ademais, apesar do AIC do modelo 3 ser maior do que do Modelo 1, foi levado em consideração primeiramente o teste de Durbin-Watson. Além disso, o modelo 1 foi o único a rejeitar o teste de normalidade dos resíduos, através do teste de Shapiro-Wilk (SW), a um nível de significância de 5% também.

Para mais, a fim de complementar o teste de normalidade, segue abaixo as análises dos resíduos do Modelo 3 (M3).

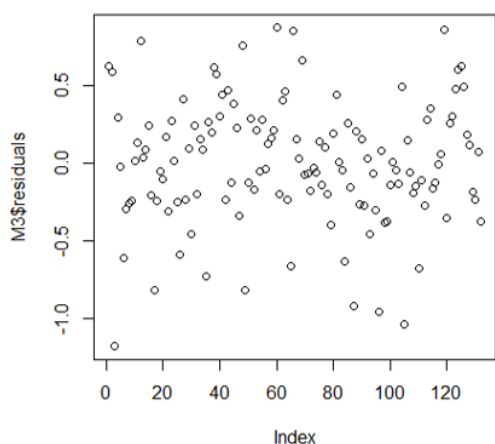


Figura 5 – Análise dos Resíduos M3

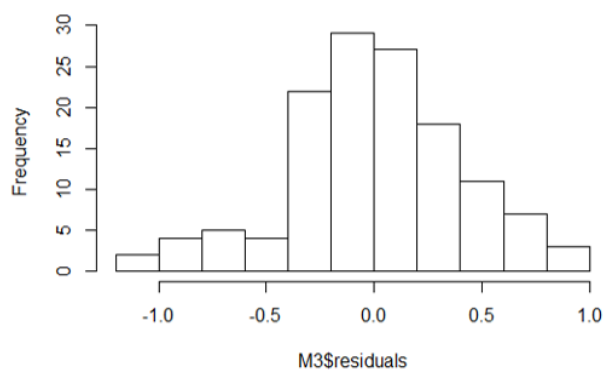


Figura 6 – Histograma dos Resíduos M3

A partir da Figura 5, é possível observar que os resíduos estão bem distribuídos, sem manifestar um padrão e, pela Figura 6, que apresentam uma assimetria em torno de zero, mas que ainda cabe uma distribuição normal ao ajuste dos dados.

Finalmente, a partir da metodologia descrita e do modelo escolhido, através do software R, obtém-se as previsões para o 1º semestre de 2022, como pode ser observado na tabela abaixo. Há também os valores reais para fins de comparação entre o de fato observado e o previsto (Real x Previsão).

Tabela 3 – Real x Previsão por mês e intervalos de confiança de 95%

Mês	Real	Previsão	IC inf	IC sup
jan/22	0,23	0,28	0,13	0,60
fev/22	0,29	0,30	0,13	0,66
mar/22	0,38	0,28	0,12	0,65
abr/22	0,34	0,33	0,14	0,81
mai/22	0,22	0,29	0,11	0,75
jun/22	0,29	0,31	0,11	0,89

É nítido que a previsão faz sentido e se adequa aos dados, principalmente quando comparada com os valores reais observados para sinistralidade do período. Todas as taxas projetadas estão próximas, salvo exceções que se aproximam de uma diferença de 30%, observadas nos meses de março e maio. Cabe mencionar os meses de fevereiro e abril que apresentaram diferenças menores que 3%.

4.5 Modelo Autorregressivo Integrado e de Médias Móveis – ARIMA

4.5.1 ACF e PACF

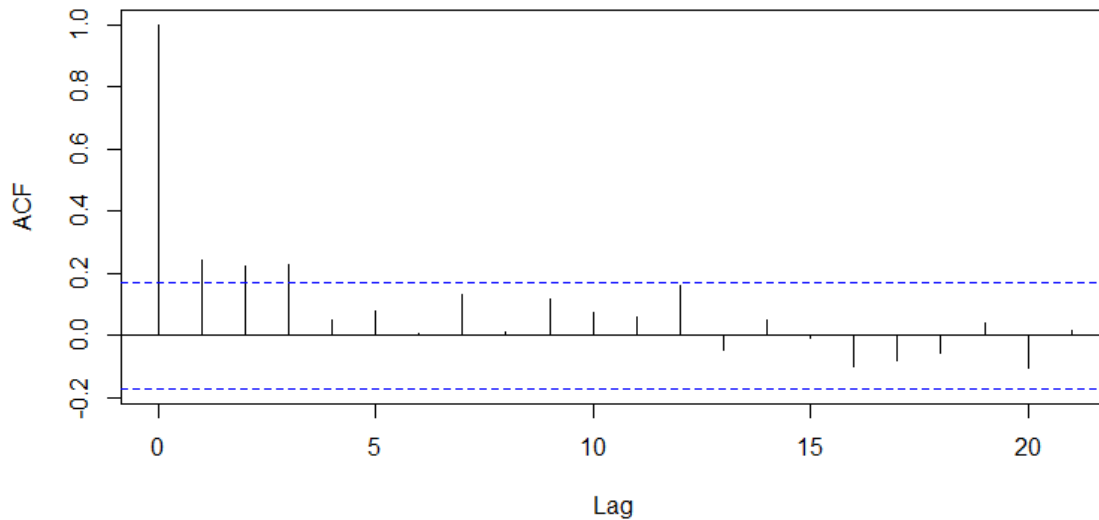


Figura 7 – ACF

Um processo estacionário tem Função de Autocorrelação (FAC) que vai para zero “rapidamente”. Isto é, a correlação em relação ao tempo passado diminui exponencialmente. Logo, em revés, um decaimento lento da FAC indica que a série não é estacionária na média e precisa passar pelo processo das diferenciações. A partir do gráfico acima, pode-se observar 3 lags fora do ponto crítico.

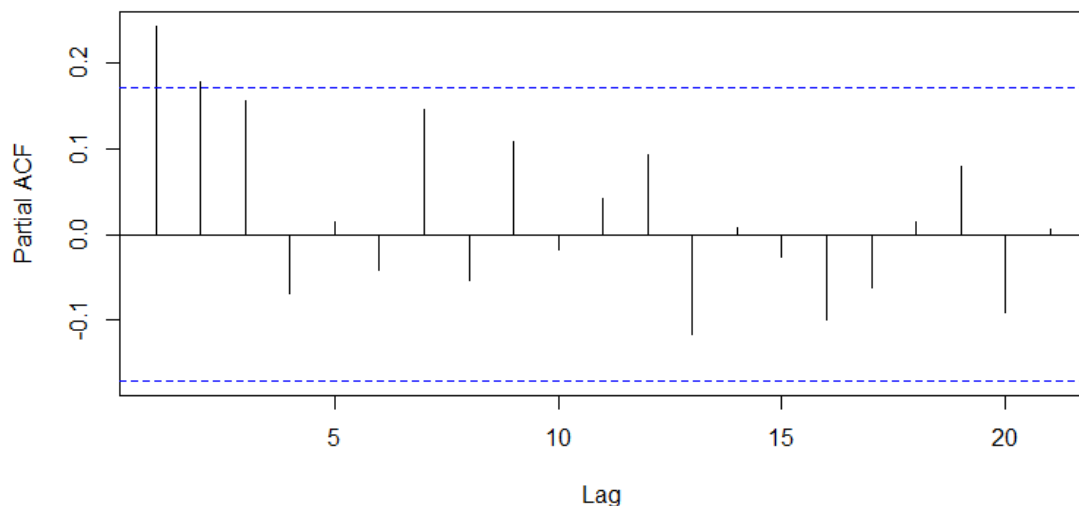


Figura 8 – PACF

No processo auto-regressivo AR, a FAC terá um decaimento exponencial ou em senoidal amortecida, se $\varphi_1 < 0$, os sinais serão alternados. A FACP tem picos significativos nos lags 1 até p, depois cai a zero. A partir do gráfico acima, pode-se observar 2 lags fora do ponto crítico.

4.5.2 Identificação do Modelo

A combinação dos modelos AR(p) e MA(q) dá origem ao modelo ARMA(p,q), todos para séries temporais estacionárias (d=0). O parâmetro p, da parte autoregressiva pode ser identificado a partir do gráfico PACF, e o q, das médias móveis, do gráfico ACF, ambos dispostos no tópico 4.5.1 ACF e PACF.

A identificação se dá pelo número de lags fora dos pontos críticos dos gráficos. Portanto, dado os 3 pontos observados no ACF e 2 no PACF, conclui-se que os parâmetros p e q são, respectivamente, 2 e 3. Logo, o modelo é ARMA(2,3) ou ARIMA(2,0,3).

4.5.3 Estimação do Modelo

Para minimizar ao máximo o número de parâmetros do modelo, foi utilizada a função `coefstest()` do pacote `lmtest` no software R. Através dela, é possível determinar quais coeficientes de fato são significantes, complementando a análise dos gráficos de autocorrelação. Nos testes foram considerados os coeficientes de maior p-valor até chegar no resultado final, um ARIMA(1,0,1).

Para complementar essa estimação do modelo, também foi utilizada a função `auto.arima()` do pacote `forecast` no software R. Essa função usa uma variação do algoritmo desenvolvido por Hyndman e Khandakar que combina testes de raiz unitária para obter um modelo ARIMA. Dentre as diversas opções de respostas que a função oferece, cabe ao interesse deste trabalho o retorno do modelo escolhido (conforme os valores de p, d e q), os termos autorregressivos AR e de média móvel MA e o intercepto.

Por fim, dentre todas as opções analisadas pelos métodos acima, foram selecionados os modelos com menores AIC, na tabela seguir.

Tabela 4 – Comparação entre Modelos ARIMA e seus respectivos AIC

Modelo	AIC
ARIMA(1,0,1)(0,0,1)[12]	134,31
ARIMA(1,0,1)(1,0,0)[12]	134,64
ARIMA(1,0,1)	136,86
ARIMA(2,0,2)	137,40

Conclui-se portanto que o melhor modelo para variável é $ARIMA(1,0,1)(0,0,1)[12]$ *with non-zero mean*. Isto é, esse resultado valida o modelo encontrado por meio da Estimação do Modelo e adiciona o segundo conjunto de parênteses, que se refere ao desenvolvimento para a parte sazonal. O valor 12 entre colchetes é o número de períodos, o que faz sentido dado que os dados observados são anuais, subdivididos entre os meses de janeiro a dezembro.

4.5.4 Diagnóstico do Modelo

A seguir, serão considerados para o diagnóstico os modelos 1 e 2, $ARIMA(1,0,1)$ e o $ARIMA(1,0,1)(0,0,1)[12]$ respectivamente, a fim de analisar o efeito da possível sazonalidade na previsão.

Para início do diagnóstico, é necessário verificar se os resíduos do modelo não são autocorrelacionados. Uma estatística utilizada para tal análise é o teste de Ljung-Box. Sua hipótese nula é que os resíduos do modelo observado são conjuntamente não correlacionados ao longo do tempo. Por conseguinte, a não independência dos resíduos seria um indicador de falha na estrutura do modelo proposto. O p-valor do modelo 1 foi de 0,4557 e do modelo 2 de 0,8859, portanto, considerando um nível de significância de 5%, não há evidências suficientes para rejeitar as hipóteses nulas, logo pode se concluir que os resíduos de ambos são conjuntamente não correlacionados.

Para complementar os testes, os correlogramas dos resíduos são expostos, abaixo nas Figuras 9 e 10. A partir deles, é possível observar que os resíduos de ambas as séries são bem distribuídos ao longo do tempo. Além disso, nos ACF, não há lags que ultrapassem as linhas de limite. Por fim, pelos histogramas, nota-se que ambos os modelos sem e com sazonalidade apresentam assimetria em torno de zero.

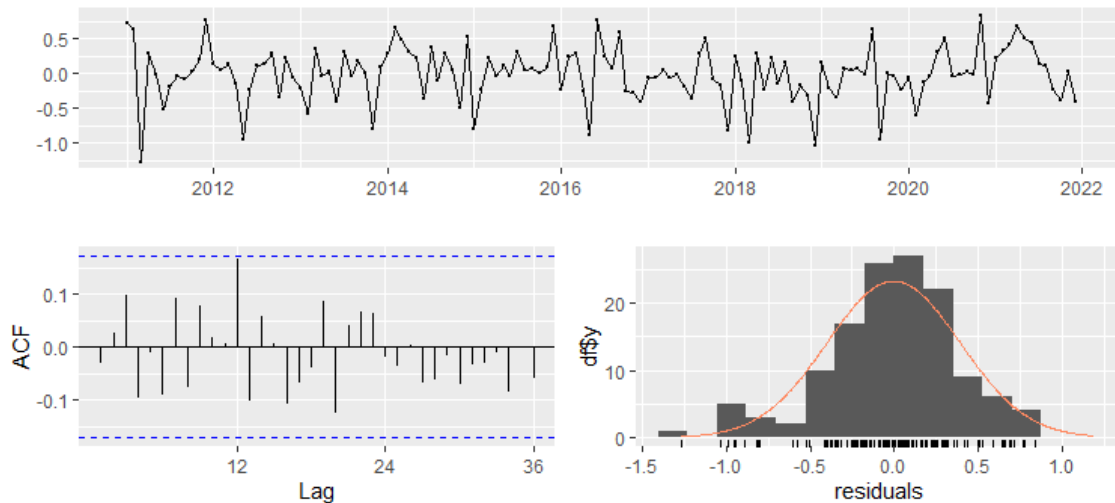


Figura 9 – Análise dos Resíduos ARIMA(1,0,1) com média diferente de zero

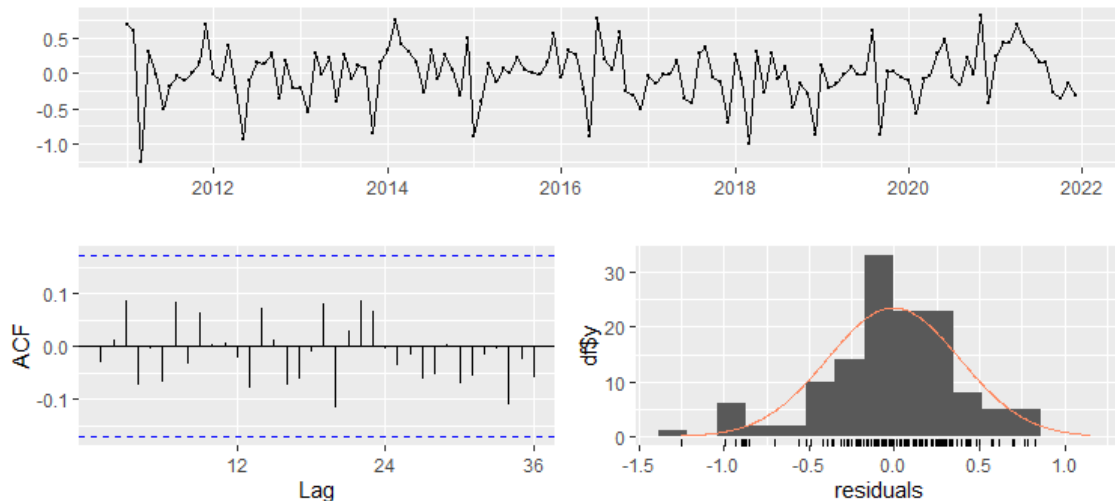


Figura 10 – Análise dos Resíduos ARIMA(1,0,1)(0,0,1)[12] com média diferente de zero

4.5.5 Previsões

Seguindo a lógica utilizada na subseção Diagnóstico do Modelo, foram feitas duas sequências de previsões, cada uma a partir de um modelo ARIMA analisado anteriormente. Finalmente, a partir da metodologia descrita e dos modelos escolhidos, através do software R, obtém-se as previsões para o 1º semestre de 2022, como pode ser observado na tabela abaixo. Novamente, há também os valores reais para fins de comparação entre o de fato observado e os previstos (Real x Previsão).

Tabela 5 – Real x Previsão por modelo e mês

Mês	Real	Previsão	
		ARIMA(1,0,1)	ARIMA(1,0,1)(0,0,1)[12]
jan/22	0,23	0,39	0,39
fev/22	0,29	0,40	0,40
mar/22	0,38	0,41	0,41
abr/22	0,34	0,41	0,41
mai/22	0,22	0,41	0,41
jun/22	0,29	0,41	0,41

Para mais, cabe uma projeção da série temporal com intervalos de confiança. Para o intervalo estimado onde a média tem uma dada probabilidade de ocorrer, foi usado níveis de 65% (representado pela porção cinza clara do gráfico) e 80% (cinza escuro).

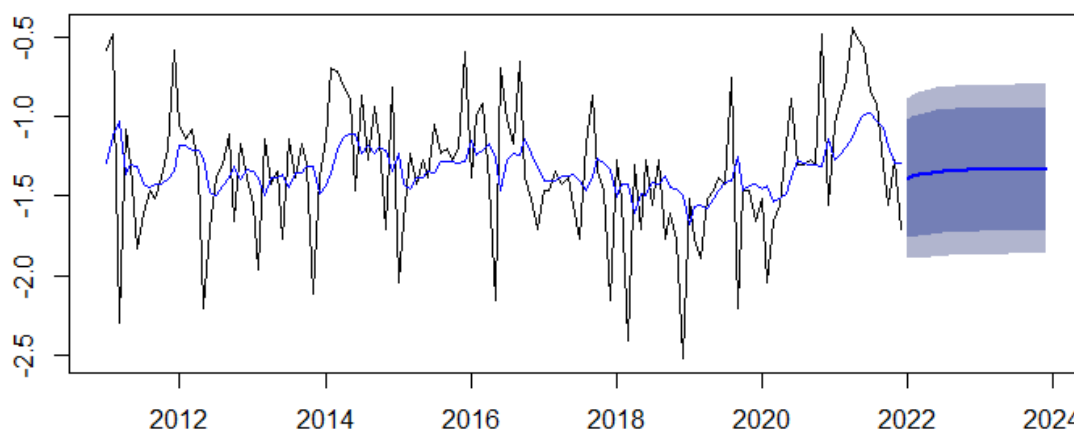


Figura 11 – Projeção da série pelo modelo ARIMA(1,0,1) com Intervalos de Confiança

Primeiramente pela projeção do ARIMA(1,0,1), pode-se observar, a partir da Figura 11 e valores da Tabela 5, que esse modelo projeta uma série de valores de crescimento constante, sem variações negativas. O ponto problemático dessa projeção é que não leva em consideração os altos e baixos que estão presentes na série observada realmente. A linha azul representa os valores ajustados do modelo para os dados observados. Logo, nota-se que essa linha projetada de sentido crescente não segue as tendências nítidas dos dados reais.

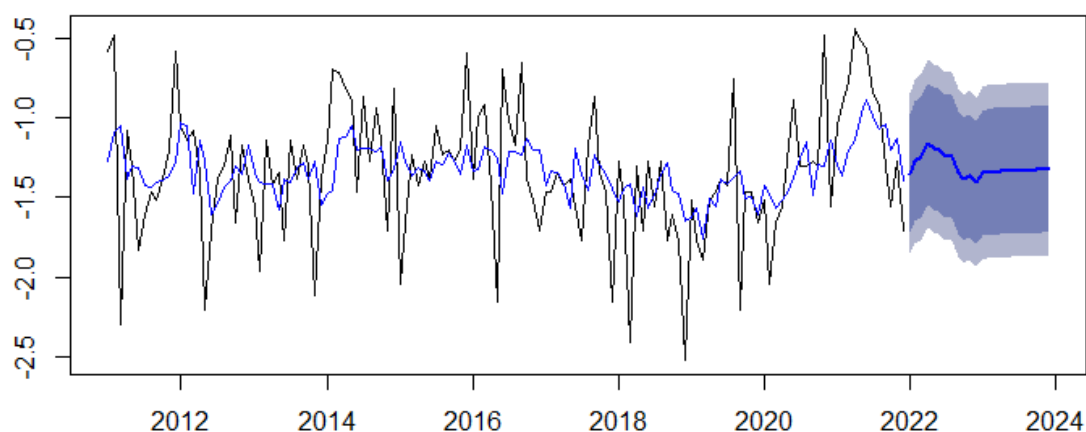


Figura 12 – Projeção da série pelo modelo ARIMA(1,0,1)(0,0,1)[12] com Intervalos de Confiança

Em revés a anterior, pode-se observar que a projeção do ARIMA(1,0,1)(0,0,1)[12], a partir da Figura 12, segue as variações presentes nos dados observados e na série, logo não é uma constante crescente para os dados futuros. Contudo, vale ressaltar que como a previsão será para os primeiros seis meses do ano de 2022, essa diferença nas projeções não tem muito impacto, porque pelo gráfico as quedas parecem começar justamente nos últimos dois trimestres de 2022, o que é confirmado pelas previsões da Tabela 5.

Outrossim, os valores entre os dois modelos de fato não se diferem significativamente. As diferenças entre previsões e real variam entre 7% a aproximadamente 80%, a depender do mês de competência. Nota-se que os valores reais oscilam e isso é esperado dado que essa taxa depende de diversos fatores, tais como o período observado e a seguradora em questão e conseqüentemente seus tipos de precificação e subscrição, por exemplo. Porém, no geral, as previsões fazem sentido, principalmente após a crise do Covid-19 e o aumento das taxas em 2021.

4.6 Comparação entre os Modelos de Previsão propostos

Para fins de comparação entre os modelos propostos nesse trabalho para previsão da sinistralidade, foi escolhida a métrica estatística dos Erros Quadráticos Médios (EQMs). Quanto menor o EQM, maior similaridade entre o real e a previsão.

Tabela 6 – Comparação entre os Erros Quadráticos Médios dos modelos de previsão

Estatística	Regressão	ARIMA(1,0,1)	ARIMA(1,0,1)(0,0,1)[12]
EQM	0,00283	0,01629	0,01595

A partir das estatísticas apresentadas na Tabela 6, o modelo com menor EQM é o de regressão. Conclui-se, portanto, que esse modelo melhor previu o 1º semestre de 2022 de acordo com a série temporal montada com dados de janeiro de 2011 a dezembro de 2021. Vale ressaltar a proximidade entre os EQMs do ARIMA(1,0,1) e o ARIMA(1,0,1)(0,0,1)[12], o que era esperado dado que as projeções foram semelhantes, como observado na subseção 4.5.5 Previsões.

5. Considerações Finais

Este trabalho realizou três modelos diferentes de previsão da Sinistralidade, para o 1º semestre de 2022, a partir dos dados observados entre janeiro de 2011 a dezembro de 2021. Isto é, foram coletadas informações de 132 meses ou 11 anos completos. O primeiro modelo foi proposto a partir de um ajuste de regressão linear simples em séries temporais, adaptado ao melhor modelo possível dentre as opções testadas. As outras duas previsões foram feitas a partir dos Modelos de Box & Jenkins ou ARIMA, diferenciando-se apenas entre a forma de estimação do modelo e, conseqüentemente, seus ajustes divergentes.

No geral, todas as previsões se adequam aos dados e estão no intervalo de confiança. Contudo, é nítido que o método de regressão projetou melhor a taxa visto que não assumiu um crescimento constante para o 1º semestre de 2022. Para mais, suas diferenças entre previsão e real foram consideravelmente melhores, como supracitado no Tópico 4.6 Comparação entre os Modelos de Previsão propostos.

Para possíveis trabalhos futuros, sugere-se o estudo da aplicabilidade de outros modelos de previsão nos dados de seguro do ramo vida individual. Além disso, cabe o uso dos modelos propostos nesse trabalho para outros tipos de seguro, como de vida em grupo, que possui outras especificidades. Há também a possibilidade de uma replicação desse estudo sem considerar os dados referentes a 2021, ano com maiores taxas de sinistralidade devido a Covid-19, a fim de analisar se de fato haveria impacto da pandemia nas previsões.

Finalmente, acredita-se que os objetivos do trabalho foram atingidos. A partir dos dados e métodos escolhidos, foram calculadas três previsões diferentes para a sinistralidade e foi possível analisar suas adequações comparando com os valores reais observados no 1º semestre de 2022.

6. Referências Bibliográficas

Antunes, José Leopoldo Ferreira; Cardoso, Maria Regina Alves. **Uso da análise de séries temporais em estudos epidemiológicos.** Epidemiol. Serv. Saúde, Brasília, 24(3):565-576, jul-set 2015.

Box, G.E.; Jenkins, G.M. **Time Series Analysis: Forecasting and Control.** Prentice Hall, Englewood Cliffs 1994.

Circular SUSEP N° 302, de 19 de setembro de 2005. Dispõe sobre as regras complementares de funcionamento e os critérios para operação das coberturas de risco oferecidas em plano de seguro de pessoas, e dá outras providências. Disponível em: <<http://www.susep.gov.br/textos/circ302.pdf>>.

Circular SUSEP N° 395, de 3 de dezembro de 2009. Estabelece a codificação dos ramos de seguro e dispõe sobre a classificação das coberturas contidas em planos de seguro, para fins de contabilização. Disponível em: <<https://www2.susep.gov.br/safe/scripts/bnweb/bnmapi.exe?router=upload/8548>>.

Correa, Bruna. **Os impactos causados nas seguradoras do ramo vida pelo aumento da sinistralidade devido a pandemia de Covid-19.** Departamento de Ciências Contábeis e Atuariais, UFRGS, 2022.

Dantas Chaves, M. E., Verola Mataveli, G. A., & Justino, R. C. (2014). **Uso da modelagem estatística para monitoramento da vegetação no Parque Nacional da Serra da Canastra, Minas Gerais.** Caderno de Geografia, 24(1), 120-132.

Coggiola, Osvaldo. **Os inícios das organizações dos trabalhadores.** AURORA ano IV número 6 – AGOSTO DE 2010; ISSN: 1982-8004.

Diniz, H.; Andrade, L. C. M.; Carvalho, A. C. P.; Andrade, M. G. **Previsão de séries temporais utilizando redes neurais artificiais e modelos de Box e Jenkins.** In: Anais do Simpósio Brasileiro de Redes Neurais. [S.l.: s.n.], 1998. p. 173–178.

Galiza, Francisco. **Seguro de Vida Individual no Brasil: o que precisa ser feito para o seu desenvolvimento?** Funenseg, 2014; 52 p.; 28 cm (Estudos sobre Seguros, n. 28).

Granger, C.W.J. and Newbold, P. **Spurious Regressions in Econometrics.** Journal of Econometrics, 2, 111-120, 1974.

Guimarães, Sérgio Rangel. **Fundamentação técnica e atuarial dos seguros vida: um estudo comparativo entre o seguro de vida individual e o seguro de vida em grupo no Brasil.** Universidade Federal do Rio Grande do Sul. Faculdade de Ciências Econômicas. Programa de Pós-Graduação em Economia. 2003.

Guimarães, S. R. **Seguros de Vida: Particularidades e mecanismos utilizados pelas seguradoras para minimizar os riscos operacionais.** ConTexto - Contabilidade em Texto, Porto Alegre, v. 2, n. 3, 2009. Disponível em:

Hanke, J.E & Reitsch A.G. **Business Forecasting.** 6th Edition, Prentice Hall, Upper Saddle River, NJ, 1998.

Huebner, S. S.; Black, K. Jr. **El seguro de vida.** Madrid: Editorial Mapfre, 1976.

Hyndman, R. J., & Khandakar, Y. **Automatic time series forecasting: The forecast package for R.** Journal of Statistical Software, 27(1), 1–22, 2008. [DOI]

Jere, S. and Moyo, E. **Modelling Epidemiological Data Using Box-Jenkins Procedure.** Open Journal of Statistics, 6, 295-302, 2016. doi: [10.4236/ojs.2016.62025](https://doi.org/10.4236/ojs.2016.62025).

Lima, Cesar de Aquino. **A história do seguro.** Revista Apólice, 2018.

Machado, Marta Alexandra Lourenço. **Modelos de previsão aplicados à optimização da gestão das actividades de um Call Center.** Universidade de Lisboa, Faculdade de Ciências Departamento de Estatística e Investigação Operacional, 2012.

Métodos de Previsão de Séries Temporais. PUC-Rio - Certificação Digital N° 0821481/CA.

Morettin, Pedro Alberto e Toloi, Clélia Maria de Castro. **Análise de séries temporais.** São Paulo: Edgard Blucher, 2006.

Mori, Flávia Tamis Medeiros; Gonçalves, Luciene Resende. **Aplicação da metodologia de séries temporais ao sistema de saúde do Brasil e dos Estados Unidos (2000-2012).** Revista Debate Econômico, v.4, n.1, jan-jun. 2016.

Nusair, Salah. **Testing the validity of purchasing power parity for asian countries during the current float.** Journal of Economic Development, v. 28, n. 2, p. 129-147, dec, 2003.

Pinheiro, Fernanda Bonfim; De Medeiros, Jussara Marques. **Origem de seguro e seguridade social.**

Reimbold, Manuel Martín Pérez, et al. **Aplicação de teste de raiz unitária as variáveis de propulsores eletromecânicos.** Vivências. Vol. 13, N.25: p.46-54, Outubro/2017.

Resolução CNSP N° 117, DE 2004. Altera e consolida as regras de funcionamento e os critérios para operação das coberturas de risco oferecidas em plano de seguro de pessoas, e dá outras providências. Disponível em: <<https://www2.susep.gov.br/safe/scripts/bnweb/bnmap.exe?router=upload/5272>>.

Resolução CNSP N° 348, DE 25 DE SETEMBRO DE 2017. Altera e consolida as regras de funcionamento e os critérios para operação da cobertura por sobrevivência oferecida em plano de seguro de pessoas e dá outras providências. Disponível em: <<https://www2.susep.gov.br/safe/scripts/bnweb/bnmap.exe?router=upload/18574#:~:text=DE%20SEGUROS%20PRIVADOS,RESOLU%C3%87%C3%83O%20CNSP%20N%C2%BA%20348%2C%20DE%2025%20DE%20SETEMBRO%20DE%202017,que%20lhe%20confere%20o%20art.>>>.

Resolução CNSP N° 439, DE 04 DE JULHO DE 2022. Dispõe sobre as características gerais para operação das coberturas de risco de seguros de pessoas. Disponível em: <<https://www2.susep.gov.br/safe/scripts/bnweb/bnmap.exe?router=upload/26145>>.

SEGURO: Uma história de 33 Séculos. Bescor, Corretora Oficial do Governo de Santa Catarina.

Shumway, Robert H.; Stoffer, David S. **Time Series Analysis and Its Applications With R Examples.** DOI: <https://doi.org/10.1007/978-3-319-52452-8>.

Vianna, Flávia Gomes. **Sinistralidade das operadoras de planos privados de assistências à saúde médico-hospitalar: determinação dos fatores associados a esse índice e seus efeitos.** Departamento de Estatística, UFMG, 2018.