



**Universidade do Estado do Rio de Janeiro**  
Centro de Tecnologia e Ciências  
Instituto de Matemática e Estatística

Felipe de Oliveira Sá

**Gestão de Risco sob a Ótica Fuzzy em Modelo Multivariado**

Rio de Janeiro

2021

Felipe de Oliveira Sá

**Gestão de Risco sob a Ótica Fuzzy em Modelo Multivariado**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.<sup>a</sup> Dra Regina Serrão Lanzillotti

Rio de Janeiro

2021

CATALOGAÇÃO NA FONTE  
UERJ / REDE SIRIUS / BIBLIOTECA CTC-A

S111 Sá, Felipe de Oliveira.  
Gestão de risco sob a Ótica Fuzzy em modelo multivariado /  
Felipe de Oliveira Sá. – 2021.  
48 f.: il.

Orientadora: Regina Serrão Lanzillotti.  
Dissertação (Mestrado em Ciências Computacionais) – Universidade  
do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Lógica difusa - Teses. 2. Inteligência computacional – Teses. 3.  
Seguros de automóveis - Teses. I. Lanzillotti, Regina Serrão. II.  
Universidade do Estado do Rio de Janeiro. Instituto de Matemática e  
Estatística. III. Título.

CDU 004.8:368

Márcia França Ribeiro – CRB7- 3664 - Bibliotecária responsável pela elaboração da ficha catalográfica

Autorizo para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação,  
desde que citada a fonte.

---

Assinatura

---

Data

Felipe de Oliveira Sá

**Gestão de Risco sob a Ótica Fuzzy em Modelo Multivariado**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Aprovada em 28 de julho de 2021.

Banca Examinadora:

---

Prof.<sup>a</sup> Dra Regina Serrão Lanzillotti (Orientadora)  
Instituto de Matemática e Estatística – UERJ

---

Prof.<sup>a</sup> Dra. Zochil González Arenas  
Instituto de Matemática e Estatística – UERJ

---

Prof. Dr. Fernando Moraes de Oliveira  
Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro –  
IFRJ

---

Prof.<sup>a</sup> Dra. Ana Paula da Silva  
Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro –  
IFRJ

Rio de Janeiro

2021

## RESUMO

SÁ, Felipe de Oliveira. *Gestão de Risco sob a Ótica Fuzzy em Modelo Multivariado*. 2021. 47 f. Dissertação (Mestrado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

O setor de seguro é de grande importância para o mercado nacional e tem papel altamente relevante no desenvolvimento e geração de renda no país. Atualmente, o mercado de seguros brasileiro continua se desenvolvendo de maneira consistente, sendo responsável por expressiva parcela do PIB (Produto Interno Bruto) e com expectativa de crescimento, considerando que ainda nos encontramos abaixo dos níveis observados em economias mais desenvolvidas. Esse nicho no Brasil corresponde ao ramo de automóveis, um dos mais expressivos, sendo que as seguradoras são responsáveis pelo retorno do patrimônio, impondo as instituições seguradoras a aperfeiçoarem a forma de precificação de seguro. Diante de um contexto de mercado segurador altamente competitivo e ágil, propõe um Modelo Multivariado Logístico Fuzzy para classificar o risco em cinco grupos: baixíssimo, baixo, médio, alto e altíssimo. Esta opção propicia a possibilidade para visão gerencial e eficiente referente precificação de seguros de automóveis, considerando as incertezas e subjetividades do mercado consumidor. Este modelo caracteriza-se em proporcionar uma solução dinâmica compatível com a velocidade das alterações legislativas que sofrem as apólices e da eficácia ao tratamento computacional da massa de informações de magnitude expressiva. O modelo proposto de classificação de risco adapta-se a mudanças, distinguindo-se da classificação inerente aos modelos que estão sendo utilizados nas instituições de seguro

Palavras-chave: Lógica Fuzzy. Aprendizado de Máquina. Seguro de Automóveis. Risco.

## ABSTRACT

SÁ, Felipe de Oliveira. *Risk Management from the Fuzzy Perspective in a Multivariate Model*. 2021. 47 f. Dissertação (Mestrado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

The insurance industry is a very important sector for the national economy and plays an important role in the development and generation of income in Brazil. The domestic insurance market keeps expanding and currently accounts for a significant portion of the national Gross domestic product (GDP). With high growth expectations it has the potential to account for an even larger part of the GDP just like in countries with stronger, more developed economies. Due to theft and property damage the automobile sector is the most important part of the domestic insurance market which forces the insurance companies to continuously develop their pricing strategies. Based in a highly competitive and agile automobile market, this paper proposes a Fuzzy Multivariate Logistic Model for risk management that offers categorization into five risk groups: very low risk, low risk, medium risk, high risk and very high risk. This enables an efficient and manageable pricing strategy that considers the uncertainties and subjectivities of the real world. The model is flexible and efficient thus able to handle policy changes as well as process large amount of data while taking into account the limitations, subjectivities and uncertainties associated with the automobile sector. It provides quick, real-time risk classification and thus differentiates from the static classifications offered by the standard models within the insurance sector.

Keywords: Fuzzy Logic. Machine Learning. Auto Insurance. Risk.

## LISTA DE FIGURAS

Figura 1	- Linha Temporal de 1808 a 1911 . . . . .	11
Figura 2	- Linha Temporal de 1919 a 2007 . . . . .	12
Figura 3	- Distribuição percentual do Volume de Prêmios pelos Segmentos do Mercado de Seguros . . . . .	14
Figura 4	- Etapas da Metodologia . . . . .	19
Figura 5	- Conjuntos Booleanos x Conjuntos Fuzzy . . . . .	21
Figura 6	- Formatos mais comuns de funções de pertinência. . . . .	23
Figura 7	- Etapas do Sistema Lógico Fuzzy . . . . .	24
Figura 8	- Conjuntos Fuzzy com as entradas padronizadas . . . . .	25
Figura 9	- Fuzzy em seguros de terremotos . . . . .	31
Figura 10	- Fuzzy em faixa etária de condutores . . . . .	32
Figura 11	- Variáveis da amostra . . . . .	33
Figura 12	- Gráfico de Dispersão das apólices segundo Exposição, Valor do veículo e Montante de Sinistros padronizadas em unidades do desvio padrão segundo as perspectivas da Exposição e Valor do Veículo . . . . .	35
Figura 13	- Distribuição das apólices por grupo de risco segundo Exposição, Valor do Veículo e Valor médio do sinistro . . . . .	37
Figura 14	- Distribuição de Frequência dos riscos padronizados em unidades do desvio padrão segundo faixas de risco . . . . .	38
Figura 15	- Análise Linear Discriminante para cinco agrupamentos em função das Componentes Principais, variáveis PC-1 e PC-2 . . . . .	39

## LISTA DE TABELAS

Tabela 1	- Receitas Anuais do Mercado de Seguro . . . . .	13
Tabela 2	- Evolução da frota de veículos do Brasil . . . . .	15
Tabela 3	- Regras de Inferência Fuzzy . . . . .	26
Tabela 4	- Tamanho das sementes segundo os atributos: saudáveis e anômalas . . . . .	27
Tabela 5	- Resumo das estatísticas gerais . . . . .	35
Tabela 6	- Distribuição das apólices entre os grupos de risco para Modelo Multivariado Logístico Fuzzy e o Modelo de Análise Linear Discriminante . . . . .	40

## **LISTA DE ABREVIATURAS E SIGLAS**

ANS	Agência Nacional de Saúde Suplementar
CNT	Confederação Nacional de Transporte
CNeg	Confederação Nacional das Empresas de Seguros Gerais, Previdência Privada e Vida, Saúde Suplementar e Capitalização
COGET	Coordenação de Apoio à Gestão Estratégica
FIFA	Federação Internacional de Futebol
PIB	Produto interno Bruto
SUSEP	Superintendência de Seguros Privados

## SUMÁRIO

	<b>INTRODUÇÃO</b>	9
1	<b>HISTÓRIA DO MERCADO DE SEGUROS BRASILEIRO</b>	10
1.1	<b>Mercado de Seguros no Brasil</b>	10
1.2	<b>Seguro de Veículo no Brasil</b>	14
2	<b>MOTIVAÇÃO</b>	16
3	<b>OBJETIVOS</b>	17
3.1	<b>Objetivos Específicos</b>	17
4	<b>METODOLOGIA</b>	18
4.1	<b>Etapas da Metodologia</b>	18
4.2	<b>Teoria dos Conjuntos Fuzzy</b>	19
4.3	<b>Funções de Pertinência</b>	22
4.4	<b>Sistema Fuzzy</b>	24
4.5	<b>Princípio de Extensão Fuzzy</b>	26
4.6	<b>Distância Euclidiana</b>	27
4.7	<b>Distância de Mahalanobis</b>	27
4.8	<b>Análise Linear Discriminante</b>	28
4.9	<b>Teste do Qui-Quadrado de Pearson</b>	28
5	<b>LÓGICA FUZZY EM SEGUROS</b>	30
6	<b>ANÁLISE EXPLORATÓRIA DOS DADOS</b>	33
6.1	<b>Análise de dados gerais</b>	34
7	<b>RESULTADOS</b>	37
7.1	<b>Validação do Modelo</b>	39
	<b>CONSIDERAÇÕES FINAIS</b>	41

## INTRODUÇÃO

O Brasil vivencia uma grande evolução na oferta de produtos de seguros com vasto e contínuo aumento das receitas das seguradoras, incidindo na gradativa participação no Produto Interno Bruto (PIB) do país. O desempenho das atividades das seguradoras tem contribuído para a eficiência do Sistema Financeiro Nacional, pois o mercado segurador contribui para a redução dos custos de transações, geração de liquidez, fomenta a economia, gerencia riscos e incentiva a captação da poupança a longo prazo.

Em 2018, o faturamento dos mercados de seguros, supervisionados pela Superintendência de Seguros Privados (SUSEP), atingiu R\$ 245,6 bilhões, pois proporcionou significativa colaboração para a formulação da poupança nacional e o desenvolvimento econômico do país, tendo participação de 3,6% no PIB em 2019 e possível crescimento na ordem de 6,0% a 10,0% (COGET, 2019).

O segmento de seguro de automóveis é um dos que apresentam maior volume de prêmios diretos, alcançando 32,6% em prêmios de seguros, não incluindo neste total os planos de saúde. O prêmio corresponde a importância paga pelo segurado ou estipulante (proponente) à seguradora para que esta assuma o risco a que o segurado está exposto (CIRCULAR SUSEP 306/05, 17 de novembro de 2005). O cálculo dos prêmios é cuidadosamente analisado a fim de evitar perdas de liquidez da seguradora sem que afete a capacidade de arcar com os riscos assumidos e a competitividade de mercado, mensurar o prêmio sem que sem a ocorrência de prejuízo. Essa questão é a principal meta deste trabalho, qual o real risco a ser assumido pela instituição segundo as características da apólice. Nesse contexto, o mercado de seguros e resseguros é desafiado a oferecer respostas efetivas sem descuidar da rentabilidade, assim surge a proposta de implantação de algoritmos computacionais com o uso de Algoritmos Fuzzy na linguagem computacional Python a estrutura de dados de seguro de veículos no sentido do reconhecimento de padrões de risco a apólices.

## 1 HISTÓRIA DO MERCADO DE SEGUROS BRASILEIRO

A história do Mercado de seguros no Brasil teve início com a abertura dos portos ao comércio internacional, em 1808. A partir deste fato, a instituição *Companhia de Seguros BOA-FÉ* foi criada com o objetivo de operar com seguros marítimos, estreando a atividade seguradora brasileira. A seguir estão os fatos importantes do mercado segurador na Figura 1 e 2.

Linha do Tempo com Fatos Históricos do Seguro no Brasil, Figura 1, mostra a evolução do mercado de seguros no Brasil, iniciado com a criação de instituições, mas havia a ausência do Estado em regular as atividades, mas posteriormente houve a criação das entidades reguladoras, permitindo maior credibilidade e confiança nas atividades exercidas.

Esta confiabilidade contribuiu para o pleno funcionamento das instituições e propiciou a instalação de empresas internacionais no Brasil, porém submetidas ao decreto da aplicação de reservas técnicas no país.

O aquecimento do mercado de seguros estendeu-se a , uma abordagem mais social, a partir de seguros de vida, seguros de acidentes de trabalho, pensão as famílias e aposentadorias. Logicamente, cada vez mais é necessário o fortalecimento e a criação de instituições fiscalizadoras.

Entre meados da década de 1970 e o final da década seguinte, o mercado de seguros se encontrou em estagnação, mas voltou a crescer de forma mais acentuada a partir da década de 1990 em função da concessão do governo às seguradoras à liberdade de operacionalização e também a estabilização monetária de 1994 (BCB, 2020).

O volumoso crescimento dos planos de saúde conduziu a instituição da Agência Nacional de Saúde Suplementar, permitindo um olhar mais concentrado a este ramo. Outro grande fato que proporcionou o crescimento dos seguros foram as grandes obras de infraestrutura geradas pelo Programa de Aceleração do Crescimento, as da Copa Mundial de Futebol promovida pela Federação Internacional de Futebol (FIFA), em 2014 e as dos Jogos Olímpicos na cidade do Rio de Janeiro, em 2016.

### 1.1 Mercado de Seguros no Brasil

Mesmo considerando toda a história dos seguros no Brasil, o país ocupa a 12<sup>a</sup> posição no ranking mundial em termos de emissão de prêmios totais, nos ramos: vida e não vida. Os índices de consumo per capita de seguros ainda se encontram em patamares julgados baixos em relação a países desenvolvidos, tornando-o um mercado de grandes oportunidades e de desenvolvimento do setor (SWISS RE INSTITUTE, 2018).

Estima-se a necessidade de investimento de 1,7% trilhão para viabilizar 2.663 projetos de infraestrutura que viriam a solucionar problemas de modernização do setor de transporte

Figura 1 - Linha Temporal de 1808 a 1911

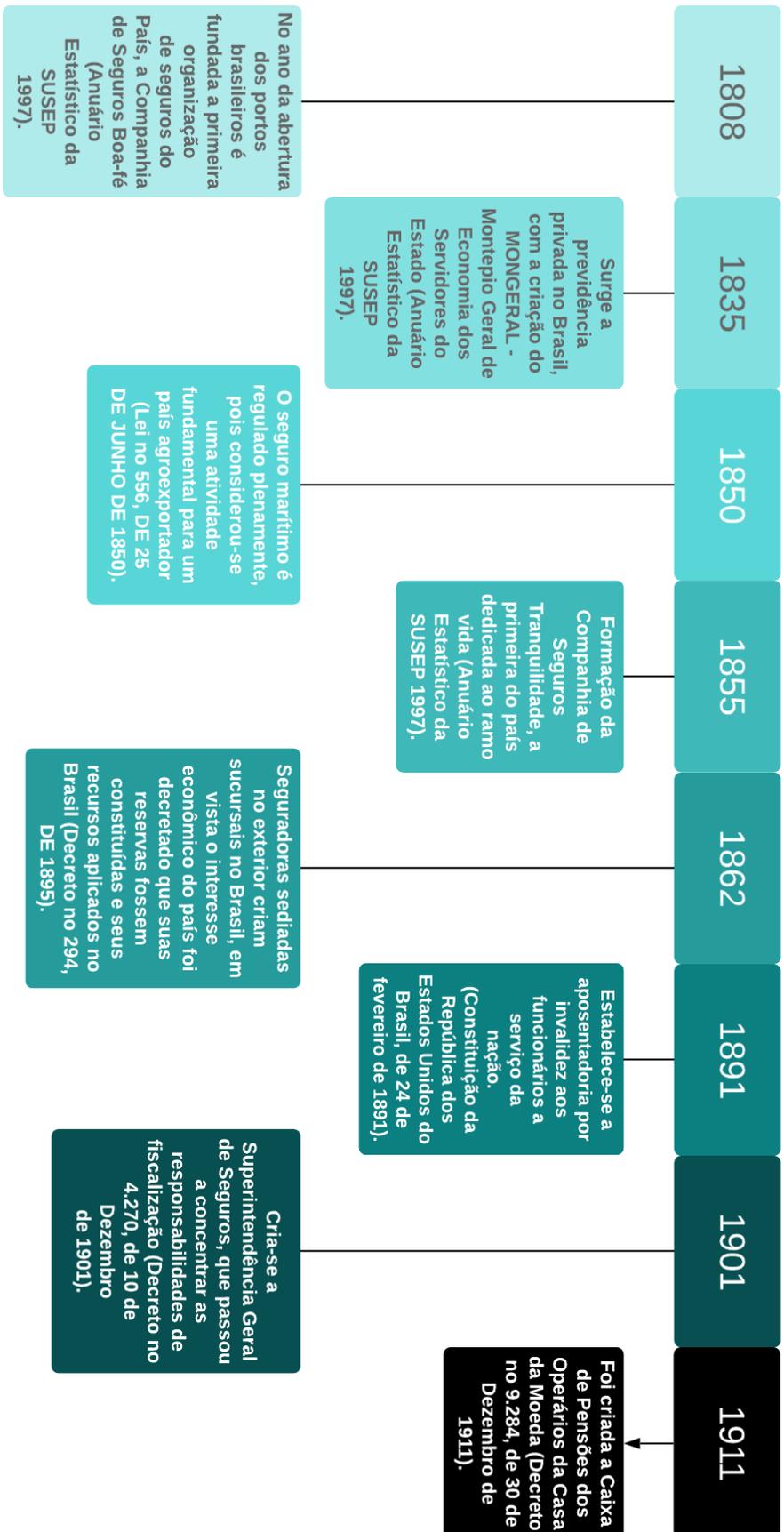
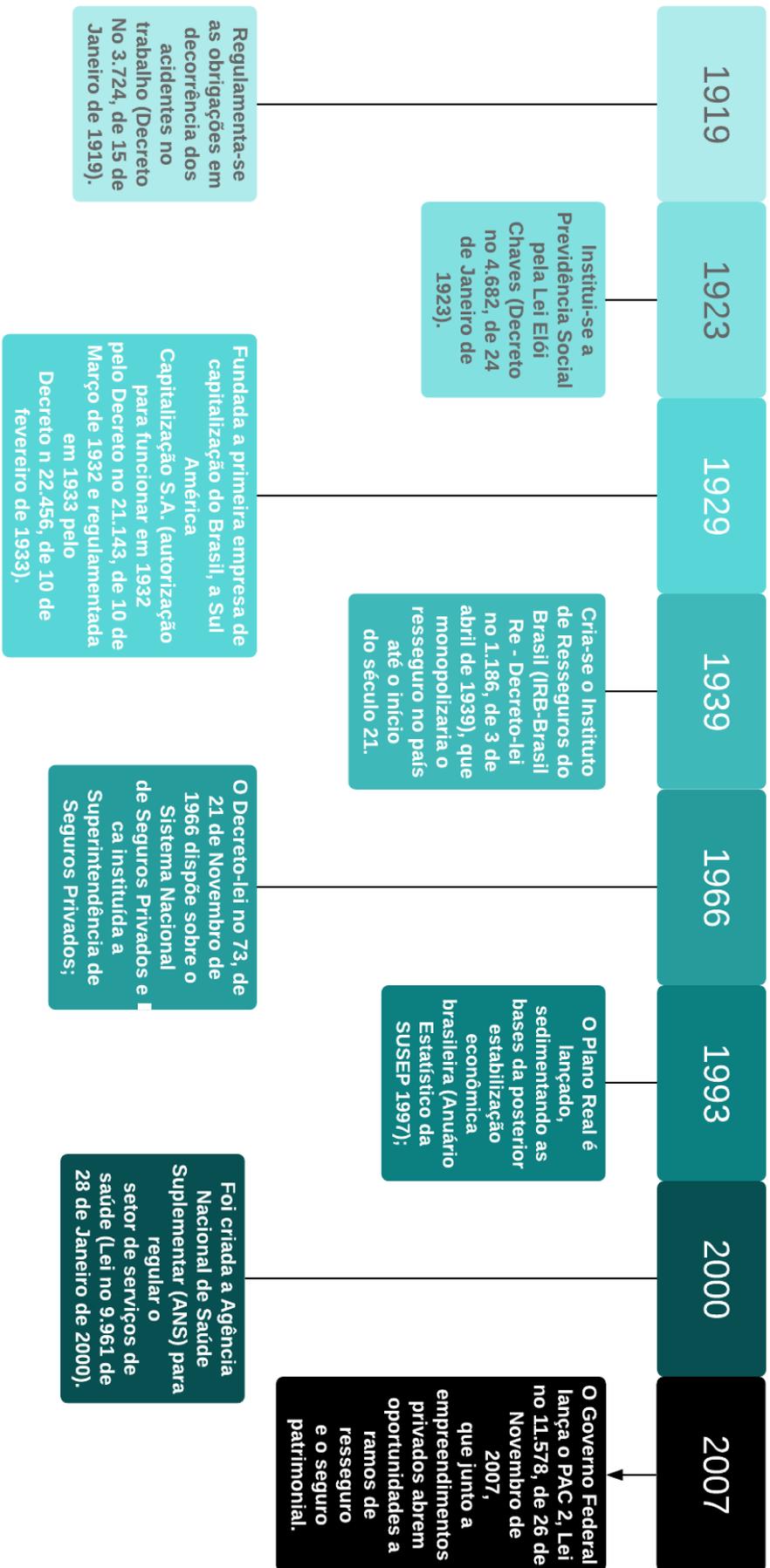


Figura 2 - Linha Temporal de 1919 a 2007



brasileiro (CNT, 2018), conseqüentemente o risco eminente da gestão desta iniciativa passará pelo mercado segurador.

A Confederação Nacional das Empresas de Seguros Gerais, Previdência Privada e Vida, Saúde Suplementar e Capitalização (CNseg) em coletiva de imprensa realizada na sede da confederação das seguradoras no Rio de Janeiro, em janeiro de 2019, possui expectativa de avanço do setor de seguros em 8,4% e 6,3% em cenários otimista e pessimista respectivamente. Seja qual for a política que o governo implementar, todos os segmentos de seguros serão contemplados, pois o setor faz parte da resiliência das economias em qualquer país do mundo, seja para apoiar a saúde da população, administrar a renda para a aposentadoria, bem como mitigar riscos dos projetos de infraestrutura tão necessários para o crescimento do Brasil (CORIOLANO, 2018).

Hoje, supervisionadas pela SUSEP, existem 119 sociedades seguradoras, 15 entidades abertas de previdência complementar, 17 sociedades de capitalização e 133 Resseguradoras. Estas instituições participam crescentemente) do PIB do país, Tabela 1. Nota-se que há uma tendência de crescimento a cada ano, apesar de um pequeno recuo em 2018, gerada pela maior atratividade do mercado de capitais (COGET, 2019).

Tabela 1 - Receitas Anuais do Mercado de Seguro

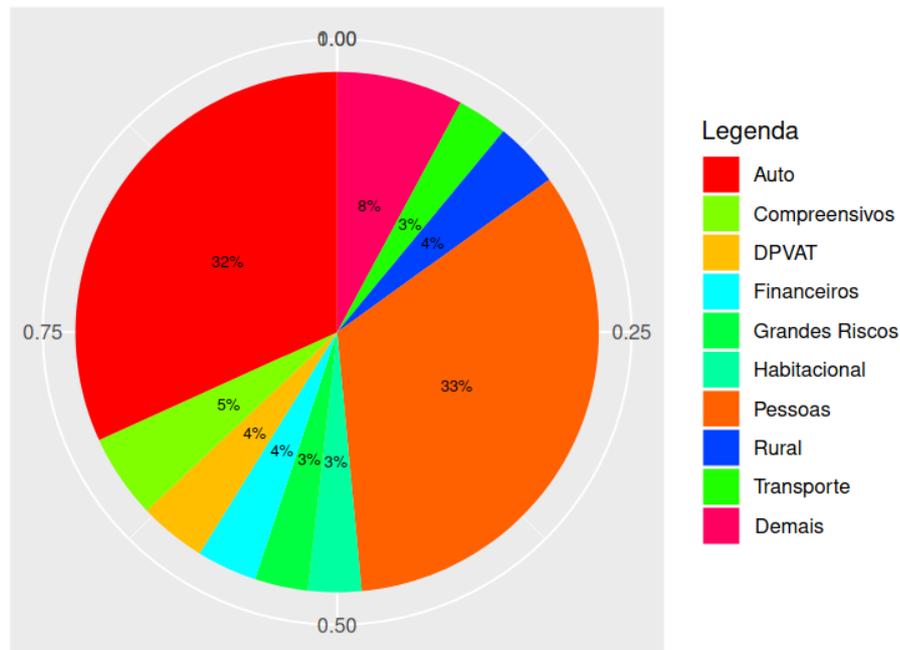
<b>Ano</b>	<b>Receitas Anuais R\$ mil</b>	<b>% PIB</b>
2003	44.552.159	2,59%
2004	52.152.043	2,66%
2005	57.210.792	2,64%
2006	63.978.067	2,66%
2007	74.186.935	2,73%
2008	85.125.692	2,74%
2009	95.269.780	2,86%
2010	111.229.059	2,86%
2011	129.423.540	2,96%
2012	157.016.628	3,26%
2013	178.013.320	3,34%
2014	198.570.647	3,44%
2015	219.027.607	3,65%
2016	239.339.365	3,82%
2017	247.245.679	3,77%
2018	245.620.047	3,60%

Fonte: (COGET, 2019)

Na Figura 3 encontra-se o Gráfico de Setores com a distribuição percentual do volume de prêmios diretos em função dos principais segmentos do mercado de seguros em 2018. Os seguros de pessoas e de automóveis tem participações percentuais semelhantes em volumes de

prêmios. Fica evidente a relevância do ramo relacionados a veículos, o que pode ser consequência de diversos fatores que foram abordados na próxima seção.

Figura 3 - Distribuição percentual do Volume de Prêmios pelos Segmentos do Mercado de Seguros



Fonte: (COGET, 2019)

## 1.2 Seguro de Veículo no Brasil

Um dos fatores que possui relação direta com a quantidade de prêmios do seguro de veículos corresponde ao aumento de veículos no Brasil de 45,37 milhões para 104,78 milhões entre os anos de 2006 e 2019, Tabela 2. O crescente aumento da frota de automóveis particulares pode ser explicado pela péssima qualidade do transporte público e o incentivo ao consumo individual nas metrópoles, estimulando ao crescimento médio anual de 6,90%, de acordo com Departamento Nacional de Trânsito, Ministério das Cidades.

A imensa circulação de veículos e os congestionamentos causam retenção de fluxo e provocam demora aos destinos dos condutores e o volume expressivo de veículos tem causado acidentes e a incidência de furtos e roubos principalmente nas metrópoles. O Anuário Brasileiro de Segurança Pública e o Fórum Brasileiro de Segurança Pública apontam roubos e furtos de 543.991 veículos em todo território nacional em 2017 (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2018).

O Departamento de Polícia Rodoviária Federal do Ministério da Justiça e Segurança Pública aponta que em 2020, 22.533 veículos envolveram-se em acidentes, causados princi-

palmente por falta de atenção à condução, enfatizando que parte dos acidentes é causada por negligência, despreparo e hábitos inadequados dos motoristas, mas também podem estar associados a condições climáticas e falhas mecânicas (POLÍCIA RODOVIÁRIA FEDERAL, 2020).

Tabela 2 - Evolução da frota de veículos do Brasil

<b>Ano</b>	<b>Frota</b>	<b>Crescimento</b>
2006	45.372.640	-
2007	49.644.025	9,4%
2008	54.506.661	9,8%
2009	59.361.642	8,9%
2010	64.817.974	9,2%
2011	70.543.535	8,8%
2012	76.137.191	7,9%
2013	81.600.729	7,2%
2014	86.700.490	6,3%
2015	90.686.936	4,6%
2016	93.867.016	3,5%
2017	97.091.956	3,4%
2018	100.746.553	3,8%
2019	104.784.375	4,0%

Fonte: (DENATRAN, 2020)

Considerando tal realidade, é possível reconhecer as determinantes que levam a população brasileira a assegurar seus automóveis e seu incremento ao longo dos anos vem aquecendo o mercado de seguros em nível nacional.

O seguro de veículos já foi considerado o de maior relevância dentro do mercado de seguros no Brasil, sendo que em 2013, tendo índice de 35% em relação aos principais segmentos, seguido de 31% referente a pessoas, sendo que atualmente estes percentuais apontam para 32% e 33%, respectivamente (COGET, 2019).

O mercado tem sido relacionada acionado aos produtos de previdência e a competição entre as seguradoras de automóveis provocando o barateamento dos prêmios e, conseqüentemente, diminuição da receita, mas para muitos brasileiros, o maior risco aos bens, continua sendo o roubo e os acidentes de automóveis.

## 2 MOTIVAÇÃO

A importância do mercado segurador, em especial ao segmento de veículos no cenário econômico nacional considera as perspectivas positivas de crescimento do segmento, em que pese à crise da economia brasileira, sendo de inegável relevância os estudos que visem compreender as suas especificidades. O reconhecimento de padrões quanto a mensuração do risco assumido em função do montante de sinistros é relevante para que se pague apenas o seu próprio risco. Isso é determinante para a saúde financeira da seguradora, tanto para não cobrar valores abusivos aos segurados, pois pode vir a perder para a concorrência quanto ao valor mínimo cobrado.

Uma instituição que não possui uma política de preços definida, estudada e avaliada estará atuando sem rumo, ou vendendo muito com prejuízo, ou perdendo negócios, ou até negociando menos, sem a clara consciência de sua performance, do impacto do mercado e da condição competitiva (BERNARDI, 1996).

Atualmente, a maioria das seguradoras utilizam os modelos estatísticos inferenciais para fornecerem informações estratégicas para tomada de decisões, precificação das apólices, manter a liquidez da empresa, criação de novos produtos e aperfeiçoamento dos existentes. Uma técnica muito utilizada é a de Modelos Lineares Generalizados usada para avaliar a relação entre a variável resposta e as variáveis explicativas, sendo o ajustamento subsidiado pela distribuição da família exponencial (JONG; HELLER, 2008).

É importante salientar que os modelos estatísticos e atuariais usados pelas instituições seguradoras, muitas vezes esbarram na lentidão quanto a adaptação de novas políticas de subscrição, o que vem favorecer modelos computacionais de eficiência comprovada que se ajustem a estrutura de dados. Diante deste contexto de mercado altamente competitivo, as seguradoras buscam implementar plataformas baseadas na Inteligência Artificial alinhadas a algoritmos de estratégias que conduzam ao sucesso pelo processamento de informações em tempo real, considerando as incertezas e subjetividades do mundo real e que motivou aplicar o algoritmo computacional Fuzzy a veículos.

Diante deste contexto e de um mercado altamente competitivo e ágil, seguradoras buscam implementar plataformas baseadas em conhecimento que estejam alinhadas a algoritmos de estratégias que conduzam ao sucesso pelo processamento de informações em tempo real, considerando as incertezas e subjetividades do mundo real. Desta forma surge a proposta deste trabalho em aplicar algoritmo computacional da área chamada Lógica Fuzzy em seguros de veículos.

### 3 OBJETIVOS

O objetivo geral dessa dissertação é usar o Modelo Multivariado Logístico Fuzzy Fuzzy para o reconhecimento de padrões em apólices de seguros de veículos que propicie uma classificação em cinco faixas de risco: Baixíssimo Risco, Baixo Risco, Médio Risco, Alto Risco e Altíssimo.

#### 3.1 Objetivos Específicos

Quanto aos objetivos específicos tem-se:

- Adotar o algoritmo Fuzzy Multivariado de Métrica Logística pelo Princípio de Extensão Fuzzy para estabelecer agrupamentos e reconhecer padrões de risco;
- Verificar se existe diferença significativa entre os grupos segundo a Análise Inferencial Estatística de variáveis inerentes ao risco;
- Oportunizar bonificação para as apólices que apresentam risco reduzido;
- Possibilitar a revisão dos formulários de apólice para as políticas de subscrição.

## 4 METODOLOGIA

Esta pesquisa utilizou como variáveis de entrada do Modelo Multivariado Logístico Fuzzy referentes a seguros de veículos, que foram normalizadas: Exposição, Importância Segurada e Montante de Sinistros. O Modelo foi operacionalizado segundo o Princípio de Extensão Fuzzy pela Distância de Mahalanobis, transformando a função logística clássica em Conjuntos Fuzzy, cujos valores de pertinência encontram-se no intervalo  $[0,1]$ . Estes valores foram divididos em cinco grupos em função dos percentis que delimitaram seus intervalos e foram designados os termos linguísticos Fuzzy: Baixíssimo Risco, Baixo Risco, Médio Risco, Alto Risco e Altíssimo.

Optou-se pela Distância de Mahalanobis pois considera as correlações entre as variáveis, corrigindo algumas limitações da Distância Euclidiana. Os algoritmos da construção do modelo encontram-se no Anexo A, implementados na linguagem Python com o uso dos pacotes:

- Adotar o algoritmo Fuzzy Multivariado de Métrica Logística pelo Princípio de Extensão Fuzzy para estabelecer agrupamentos e reconhecer padrões de risco;
- Numpy (Computação científica com Python);
- Pandas (Análise e manipulação de dados);
- Scipy (Código aberto para matemática, ciências e engenharia);
- Sklearn (Análise preditiva de dados);
- Matplotlib (Criação de visualizações).

O modelo foi construído com auxílio desses pacotes e apresenta uma complexidade de  $O(N)$ . Complexidade de algoritmo é a quantidade de trabalho necessário para executar uma tarefa, logo o tempo de execução do modelo é linearmente proporcional a quantidade de registros que possui no base de dados.

### 4.1 Etapas da Metodologia

A Lógica Fuzzy nos permite converter e incorporar conhecimentos qualitativos empíricos em sistemas lógicos capazes de executar correspondências e reconhecer padrões com base no conhecimento dos dados existentes.

Figura 4 - Etapas da Metodologia



Fonte: O autor, 2020.

1. Normalização das variáveis: Inicialmente, as variáveis: Exposição, Importância Segurada e Montante de Sinistros foram normalizadas em unidades do desvio padrão. Estes valores foram utilizados como entradas para o Modelo Multivariado Logístico Fuzzy.
2. Cálculo da Distância de Mahalanobis: A partir das variáveis de entradas normalizadas encontrou-se o centroide da distribuição pela Distância de Mahalanobis, possibilitando mensurar o distanciamento do centróide.
3. Aplicação na Equação Logística: As distâncias permitiram aplicar o Princípio de Extensão para usar a Função Logística Fuzzy, resultando nos valores de pertinência de cada apólice.
4. Classificação pelos percentis; Os valores de pertinência permitiram classificar as apólices em cinco faixas de risco segundo cortes percentílicos, discriminados: Até o Percentil 20, Acima do Percentil 20 até o Percentil 40, Acima do Percentil 40 até o Percentil 60, Acima do Percentil 60 até o Percentil 80, Acima do Percentil 80 até o Percentil 100, adotando os termos linguísticos: Baixíssimo risco, Baixo risco, Médio risco, Alto risco e Altíssimo risco.
5. Validação pela LDA: Essa proposta de reconhecimento de padrões sob a ótica Fuzzy foi validada segundo o teste Qui-quadrado segundo a Tabela de Continência Categórica dos termos linguísticos, onde cada célula agrega a frequência por categoria de apólice segundo as classificações dos modelos Fuzzy e da Análise Linear Discriminante.

## 4.2 Teoria dos Conjuntos Fuzzy

A compreensão da Teoria dos Conjuntos Fuzzy partiu da Teoria dos Conjuntos Bivalentes Clássicos, bastante utilizada pela maioria dos modelos propostos, sendo que a primeira possui seus limites bem definidos: se pertence-se ou não a um determinado grupo, verdadeiro ou falso, os conjuntos são considerados determinísticos e precisos. Na segunda teoria, existe a possibilidade de se pertencer aos dois conjuntos simultaneamente (ZIMMERMANN, 2001). Essa precisão encontra uma barreira, pois:

À medida que a complexidade de um sistema aumenta, nossa capacidade de fazer afirmações precisas e significativas sobre seu comportamento diminui até que

seja atingido um limite além do qual precisão e significância (ou relevância) se tornam características quase mutuamente exclusivas (ZADEH, 1973).

A maioria dos modelos adota a precisão de pertencimento, mas nem sempre esta adoção pode ser adotada e tendem a fracassar, pois os fenômenos no mundo real são ambíguos, incertos, vagos e frequentemente subjetivos.

Os conceitos dos Conjuntos Fuzzy podem ser vistos como uma generalização dos Conjuntos Bivalentes Clássicos, possuindo transições graduais definidas pelo grau de pertinência,  $\mu(x)$ , variando entre zero e 1, exclusão e pertinência completa, respectivamente. (BARROS; BASSANEZI, 2006).

Ao modelar aspectos imprecisos a partir de uma perspectiva alternativa de estrutura que substitui a lei do meio excludente de Aristóteles pela lógica de Bertrand Russel, surge a Lógica Fuzzy, em que uma afirmativa ambígua pode ter valores em uma escala intervalar entre 0 e 1, valores considerados de pertinência (LANZILLOTI; LANZILLOTI, 1999).

Imagine um problema simples, em que se pretende construir um modelo que controle o aparelho de ar condicionado de um determinado local para se economizar energia. O modelo terá duas ações: ligar e desligar. Logicamente, ligar quando a temperatura está quente e desligar quando está fria. Pode-se determinar acima de  $38^{\circ}\text{C}$  uma temperatura quente, então, baseado na Lógica Bivalente Clássica, este fenômeno pode ser modelado da seguinte forma:

$$f(x) = \begin{cases} \text{Se temperatura} > 38^{\circ}\text{C}; & 1; \text{ (Ligar o aparelho);} \\ \text{Caso contrrio;} & 0; \text{ (No ligar o aparelho).} \end{cases} \quad (1)$$

O modelo funciona perfeitamente, uma vez que foram estabelecidos os limites rígidos de temperatura quente e fria, mas em qual temperatura será ligado ou o aparelho? A temperatura de  $37,9^{\circ}\text{C}$  seria considerada uma temperatura baixa, mesmo estando mais próximo de vir a ser considerada quente, que considera uma pequena alteração no valor conduzir a uma classificação alterada abruptamente.

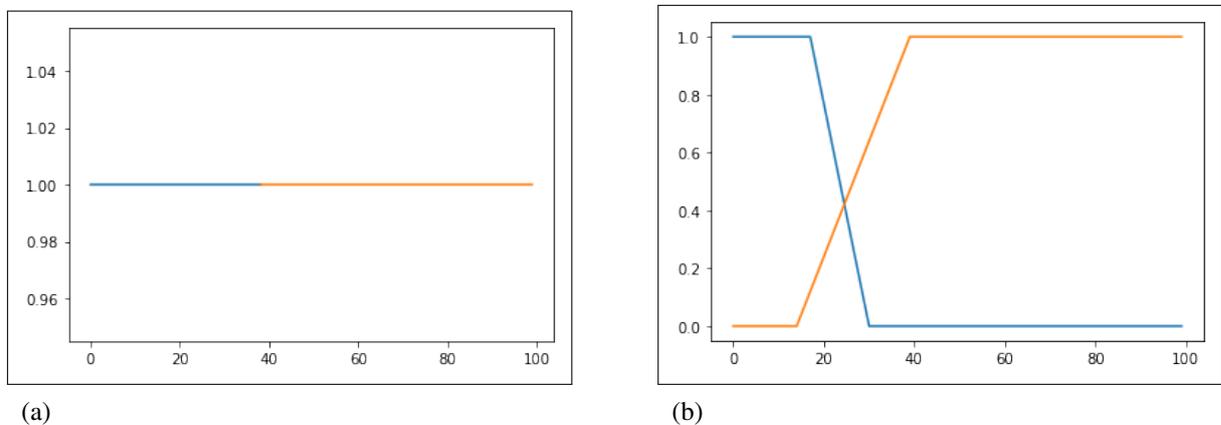
Este pequeno problema ilustra a dificuldade da lógica formal em lidar com alguns problemas reais, em razão da dificuldade em lidar com as fronteiras rígidas das classes. A teoria dos Conjuntos Fuzzy baseia-se no conceito de pertinência, que pode ser entendido como o quanto um valor pertence àquele conjunto. A modelagem estaria relacionada à intensidade à qual se pertence aquele grupo:

$$\mu_A = \begin{cases} \text{Se temperatura} > 38^{\circ}\text{C}; 1; (\text{Ligar o aparelho}); \\ \text{Se, } 18^{\circ}\text{C} < \text{temperatura} \leq 38^{\circ}\text{C}; \frac{x-18}{38-18}; (\text{Ligar o aparelho}); \\ \text{Caso contrario, } 0; (\text{No ligar o aparelho}). \end{cases} \quad (2)$$

A adoção dos Conjuntos Fuzzy permite ser discutido com que intensidade deve ser ligado o aparelho, dependendo da temperatura, pois os valores deixam de pertencer ao conjunto gradativamente. Se  $x$  está em  $37,5^{\circ}\text{C}$ , então  $\mu(x) = 0,975$ , seria como dizer que esta temperatura pertence 97,5% ao conjunto quente.

Uma frase atribuída a John Wilder Tukey, famoso matemático estadunidense, se enquadra perfeitamente aos Conjuntos Fuzzy: "É preferível uma resposta aproximada à pergunta certa, que muitas vezes é vaga, do que uma resposta exata para a pergunta errada, que sempre pode ser feita com precisão". A Lógica Clássica mensura o pertencimento de determinado valor ao conjunto, porém "frequentemente encontramos objetos no mundo real que não possuem precisamente uma definição em um certo conjunto"(ZADEH, 1965).

Figura 5 - Conjuntos Booleanos x Conjuntos Fuzzy



Legenda: Gráficos dos conjuntos das temperaturas frias e quentes, onde o eixo x representa as temperaturas e o eixo y o grau de pertinência, Azul: conjunto das temperaturas frias, Laranja: conjunto das temperaturas quentes.(a) Perspectiva da Lógica clássica. (b) Perspectiva dos Conjuntos Fuzzy.

Fonte: O autor, 2020.

Na Figura 5(a) estão os conjuntos clássicos bivalentes, pois um deles inicia e termina no início do outro. Na Figura 5(b), estão os Conjuntos Fuzzy com respectivos graus de pertinência. Exibe o patamar superior estacionário seguido da tendência de declínio. O segundo Conjunto Fuzzy indica o patamar inferior estacionário seguido da tendência crescente e finda em patamar superior estacionário. Há a área de interseção entre os conjuntos, Área Fuzzy ou Nebulosa.

### 4.3 Funções de Pertinência

As funções de pertinência,  $\mu(x)$ , definem o formato dos Conjuntos Fuzzy. Os formatos mais comuns vistos na literatura são os triangulares, os trapezoidais e os Gaussianos, Figura 6 e suas equações em equação 3, 4 e 5 respectivamente. É conveniente um estudo prévio do cenário objeto do estudo antes de se determinar o Conjuntos Fuzzy com a respectiva a função de pertinência que reflita a descrição da situação, pois se função de pertinência for inadequada, pode acarretar em saídas equivocadas da modelagem.

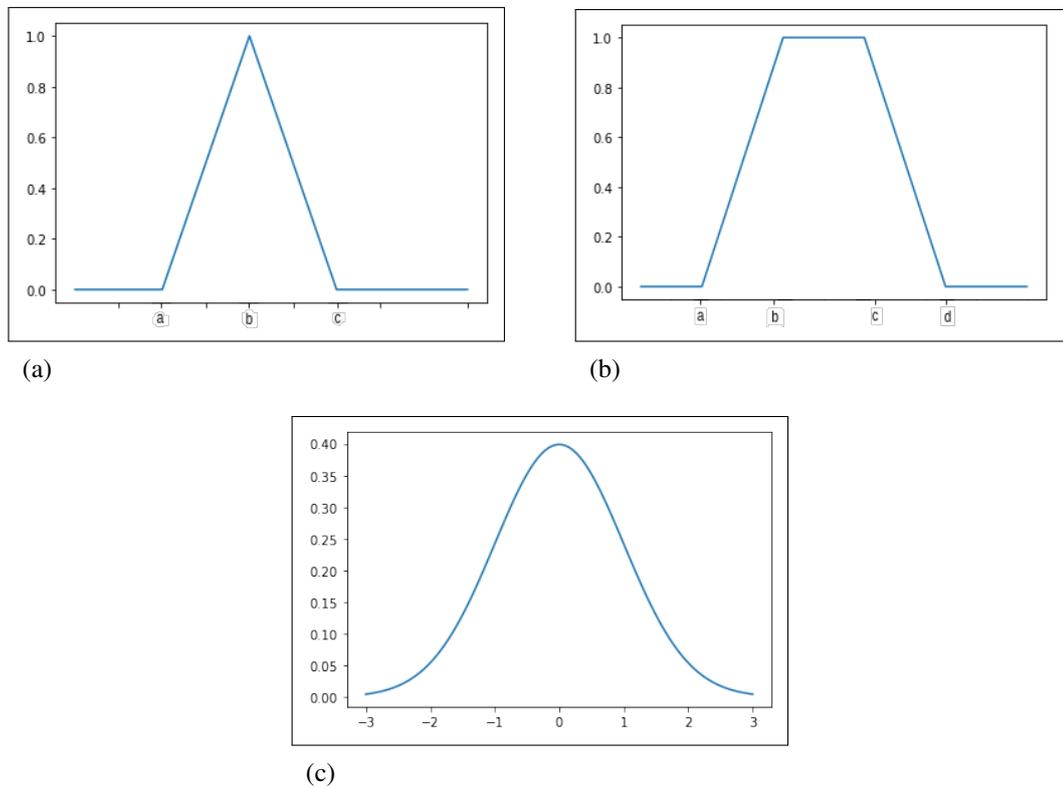
$$\mu_x = \begin{cases} 0, & \text{se } x \leq a; \\ \frac{x-a}{b-a}, & \text{se } a < x \leq b; \\ \frac{c-x}{c-b}, & \text{se } b < x \leq c; \\ 0, & \text{se } x > c. \end{cases} \quad (3)$$

$$\mu_x = \begin{cases} 0, & \text{se } x \leq a; \\ \frac{x-a}{b-a}, & \text{se } a < x \leq b; \\ 1, & \text{se } b < x \leq c; \\ \frac{d-x}{d-c}, & \text{se } c < x \leq d; \\ 0, & \text{se } x > d. \end{cases} \quad (4)$$

$$\mu_x = \begin{cases} e^{-\frac{(x-\mu)^2}{\delta^2}}; \\ 0, & \text{cc.} \end{cases} \quad (5)$$

É conveniente um estudo prévio do fenômeno antes de se determinar a função de pertinência, que reflita a descrição da situação, uma função de ertinência inadequada acarreta em saídas equivocadas da modelagem.

Figura 6 - Formatos mais comuns de funções de pertinência.



Legenda: (a) Função de pertinência triangular com parâmetros  $(a, b, c)$ , onde  $a < b < c$ ; (a) Função de pertinência trapezoidal com parâmetros  $(a, b, c, d)$ , onde  $a < b < c < d$ ; (b) Função de pertinência gaussiana,  $N(\mu; \sigma^2)$

Fonte: O autor, 2020.

#### 4.4 Sistema Fuzzy

Disposto alguns conceitos fundamentais da Lógica Fuzzy, pode-se apresentar o Sistema Lógico Fuzzy, descrevendo suas etapas, ilustradas na Figura 8. Composto por cinco etapas, inicia-se pela entrada das variáveis explicativas que o modelador julga possuir influência na variável resposta. Isso impõe a importância de se conhecer o fenômeno, analisar criticamente os dados envolvidos, inferir estatísticas e aplicar testes para confirmar e justificar a inclusão ou a não. (HOAGLIN, 2003).

Figura 7 - Etapas do Sistema Lógico Fuzzy



Fonte: O autor, 2020.

Nesta proposta, as variáveis de entrada são padronizadas segundo a diferença do valor em relação a média expressas em unidades do desvio padrão e fuzzificadas. Os termos linguísticos foram adotados para cada agrupamento obtidos segundo cortes percentílicos.

Máquina de Inferência Fuzzy corresponde a uma tabela de contingência dos Termos Linguísticos, regras antecedentes condicionais “Se...então” para reconhecer os padrões de risco nas negociações de seguro. A quantidade de regras é diretamente proporcional à quantidade de Termos Linguísticos, logo para estabelecer o número máximo de regras ( $N_{regras}$ ) usa-se a expressão:

$$N_{regras} = N_{conjuntos\ A} * N_{conjuntos\ B} \quad (6)$$

Uma abordagem muito comum nesta inferência é utilizar a pertinência mínima representado por () que possibilita encontrar o valor de pertinência entre os Termos Linguísticos. (MAMDANI; ASSILIAN, 1975).

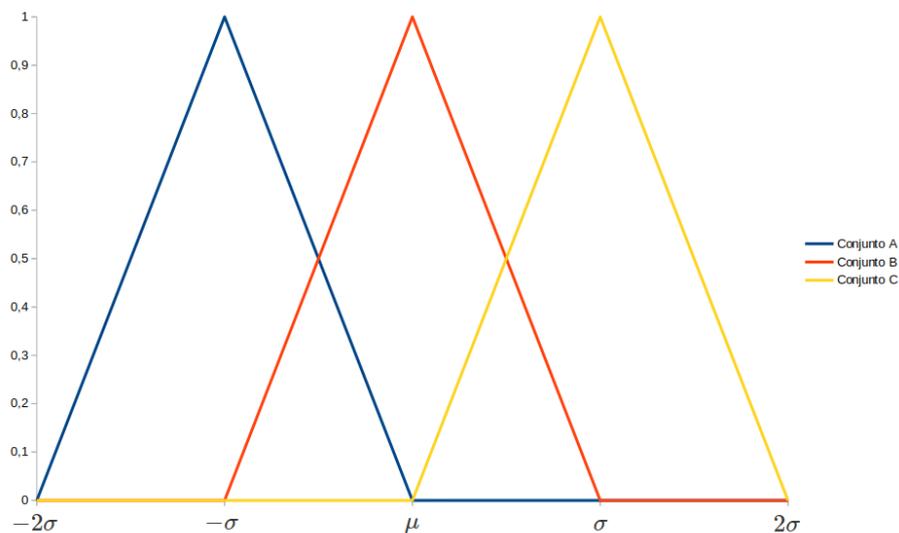
$$f(\mu_A(x), \mu_B(y)) = \mu_A(x) \wedge \mu_B(y) \quad (7)$$

A defuzzificação corresponde a reversão dos valores de pertinência referentes aos Termos Linguísticos para valores padronizados em unidades do desvio padrão e, finalmente, a conversão para a mesma métrica da entrada. Corresponde ao oposto da fuzzificação, pois transforma dados quali-quantitativos na mesma medida da entrada do sistema que pode ser encontrada a partir da proposta de (TAKAGI; SUGENO, 1985). Finalmente, chega-se a saída do sistema, referente a defuzzificação, onde se tem a variável resposta do modelo.

aplicação a um Sistema Lógico Fuzzy para precificação de seguro de automóveis, onde

pode ser utilizado as variáveis de entrada: Importância segurada, valor do veículo e Tempo de experiência do condutor. Estas variáveis podem ser padronizadas em unidades do desvio padrão processadas e fuzificadas para Termos Linguísticos, Baixo, Médio e Alto com as respectivas pertinências segundo os Conjuntos Fuzzy representados na Figura 9. Se a variável de entrada assume valor no suporte que se encontra na área de interseção de dois Conjuntos Fuzz, área nebulosa, adota-se a pertinência mínima com o operador “E” (MARQUES; BARRETO; CORTEZ, 2006).

Figura 8 - Conjuntos Fuzzy com as entradas padronizadas



Fonte: O autor, 2020.

A Máquina de Inferência Fuzzy agrega as regras Fuzzy antecedentes “Se...então” associadas às respectivas pertinências mínimas dos Termos Linguísticos aos Termos Linguísticos consequentes com a adoção da pertinência máxima dentre as pertinências mínimas aos termos linguísticos antecedentes do Sistema Lógico Fuzzy, Tabela 3 (MAMDANI; ASSILIAN, 1975).

Pode-se determinar o valor a ser assegurado para o veículo da seguinte forma: se valor do veículo e a experiência do condutor for classificado como baixo e alta com pertinências  $\mu(t) = 0,5$  e  $\mu(s) = 0,3$ , respectivamente. O Sistema Fuzzy pela defuzzificação conduz à precificação O valor do seguro pode ser encontrando através da equação:

$$V_{seguro} = \max(\mu_{valor}(v), \mu_{experiencia}(e)) * P \quad (8)$$

Onde,  $P$  é o valor da faixa de risco e multiplicado pela maior pertinência dos Termos Linguísticos.

Tabela 3 - Regras de Inferência Fuzzy

Se	Valor do veículo	E	Experiência do condutor	Então
	Baixo		Baixo	Risco médio
	Baixo		Médio	Risco baixo
	Baixo		Alto	Risco baixo
	Médio		Baixo	Risco alto
	Médio		Médio	Risco médio
	Médio		Alto	Risco baixo
	Alto		Baixo	Risco alto
	Alto		Médio	Risco alto
	Alto		Alto	Risco médio

Fonte: O autor, 2020.

#### 4.5 Princípio de Extensão Fuzzy

O Princípio de Extensão Fuzzy consiste em transformar funções dos Conjuntos Bivalentes Clássicos em Conjunto Fuzzy, alterando a Função Característica pela Função de Pertinência, dando-se liberdade em utilizar qualquer função para definir-se a Função de Pertinência. Neste trabalho, utilizou-se a Função Logística como função de pertinência ( $\mu_x$ ) e a Distância de Mahalanobis ( $D_M$ ), como métrica (MITAIM; KOSKO, 2001).

Seja  $f : U \rightarrow V$  uma função clássica. A extensão de Zadeh de  $f$  é a função  $\hat{f} : F(U) \rightarrow F(V)$  que fornece, para qualquer  $X \in F(U)$ , o Conjunto Fuzzy  $f(X) \in F(V)$ , ou seja, pretende-se encontrar uma função  $\hat{f} : F(U) \rightarrow F(V)$  que coincida com a função dos Conjuntos Clássicos (MCNEILL; THRO, 1994).

A fuzzificação do modelo proposto usou a função logística que agrega a Distância de Mahalanobis, medida que considera a variabilidade em relação ao vetor das médias e a matriz de covariância.

$$\mu_x = \frac{2}{1 + \exp(d_j(x)^2)} \quad (9)$$

Onde  $d_j(x)$  é o parâmetro da Distância Quadrática de Mahalanobis,

$$d_j^2(x) = (x - m)(Cov^{-1})(x - m) \quad (10)$$

Onde  $x$  e  $m$  indicam os vetores dos valores das observações e o vetor das médias, respectivamente (MCNEILL; THRO, 1994).

#### 4.6 Distância Euclidiana

Um bom ponto de partida para entender a Distância de Mahalanobis é conhecer a Distância Euclidiana ( $d$ ), segmento de reta entre dois pontos em um Espaço Euclidiano  $n$  – dimensional  $p_i = (p_1, p_2, \dots, p_n)$  e  $q_i = (q_1, q_2, \dots, q_n)$  obtida segundo a expressão (ALFAKIH, 2018):

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (11)$$

A título de ilustração, Tabela 4, a Distância Euclidiana pode ser aplicada para avaliar o reconhecimento de padrão de sementes sob os rótulos: saudáveis ou anômalas. O primeiro procedimento seria encontrar o centroide (média) de cada grupo que As médias para as sementes saudáveis e anômalas são 4 e 9, respectivamente. O modelo classifica uma nova semente que possui tamanho igual a 6 como saudável, mesmo possuindo o atributo do mesmo valor definido como anômalo, pois  $d_s = 2$  e  $d_a = 3$ . A Distância Euclidiana não considera o desvio em relação ao tamanho médio da semente, o que pode resultar em erros em modelos de apenas uma dimensão. A Distância de Mahalanobis considera os desvios padronizados na classificação dos grupos.

Tabela 4 - Tamanho das sementes segundo os atributos: saudáveis e anômalas

Saudáveis	Anômalas
3	6
4	8
4	12
5	10

Fonte: O autor, 2020.

#### 4.7 Distância de Mahalanobis

A Distância de Mahalanobis, Métrica Euclidiana Quadrática Ponderada tem sido usada em técnica de classificação da estatística multivariada, onde a ponderação é realizada pela inversa da Matriz de Covariância

$$d_j^2 x = (x - m)(Cov^{-1})(x - m) \quad (12)$$

Onde  $x$  e  $m$  indicam os vetores dos valores das observações e o vetor das médias, respectivamente (MAHALANOBIS, 1936).

A Distância Euclidiana pode ser considerada como um caso particular da Distância de Mahalanobis onde os desvios dos valores em relação aos centróides de cada grupo são iguais. Na exemplificação referente ao tamanho das sementes, incluindo os desvios dos tamanhos das sementes em relação aos centróides na equação da Distância Euclidiana, representada pela expressão (11), encontra-se  $d_s = 2,45$  e  $d_a = 1,16$ . Isso permite a classificação como anômala, por possuir uma distância menor.

#### 4.8 Análise Linear Discriminante

Inicialmente proposta por Ronald A. Fisher em 1936, descrito para um problema de duas classes, consiste na transformação das observações em dimensões menores, com o intuito de que os valores fossem visivelmente classificados em um plano, maximizando a distância entre os distintos e minimizando a variância entre os semelhantes (FISHER, 1936).

Posteriormente a concepção de Fisher foi generalizada como Análise Discriminante Linear multiclasse por C. R. Rao em 1948 para sua pesquisa: A utilização de múltiplas medições em problemas de classificação biológica (RAO, 1948). A Análise Linear Discriminante (LDA), técnica que utiliza o aprendizado supervisionado é frequentemente utilizada para classificação de padrões e aprendizado de máquina, reduzindo a dimensionalidade na etapa de pré-processamento dos dados, pois uma amostra  $n - dimensional$ , transforma-se a observação em um subespaço menor  $k$ , onde  $k \leq n - 1$ , mantendo as informações discriminatórias das classes, pois utiliza o comportamento das distribuições para extrair padrões lineares.

Por estas características, a LDA torna-se uma técnica de modelagem utilizada em estudos de confronto para com comprovação da sua eficácia de modelos inferenciais, sendo que no presente estudo, aferição do modelo Fuzzy Multivariado Logístico.

#### 4.9 Teste do Qui-Quadrado de Pearson

O Teste do Qui-Quadrado de Pearson ( $\chi^2$ ) julga se a diferença entre uma distribuição é relevante em relação à outra ou se tendem a se igualar, ou seja, tem a mesma distribuição de frequência. Esta opção pode comparar as distribuições em três formas. O "teste de qualidade do ajuste" ou "teste de aderência compara se a distribuição observada difere da esperada, larga-

mente usado em avaliação de modelos referentes a distribuições de probabilidades. O teste de homogeneidade compara a igualdade entre as distribuições de frequência, ao passo que o teste de independência avalia a independência das observações entre si (GREENWOOD, 1996).

$$\chi^2 = \sum_{j=1}^n \frac{(O_j - E_j)^2}{E_j}, \quad (13)$$

Onde  $O_j$  corresponde à frequência observada;  $E_j$  à frequência esperada e  $n$  é o total de observações.

A determinação dos graus de liberdade ( $gl$ ) para o teste de aderência corresponde a a diferença entre as categorias e o parâmetros da distribuição ajustada. Nas tabelas de contingências, o  $gl$  é obtido pelo produto, (linhas -1) e (colunas -1) A opção pelo nível de significância,  $p_{valor}$ , é pré-fixado em função do risco probabilístico para o resultado do teste. O confronto da Estatística Teste Qui-Quadrado Calculada e o Qui-Quadrado tabelado permite a tomada de decisão da aceitação ou rejeição da Hipótese Nula ( $H_0$ , não há diferença entre as distribuições). Se o valor Qui-Quadrado Calculado for menor ou igual ao valor crítico encontrado na tabela, aceita-se a hipótese  $H_0$ , as distribuições tendem a se igualar, caso contrário, rejeita-se a hipótese nula. (SIEGEL; CASTELLAN, 2006).

Percebe-se que a aceitação das hipóteses tem uma alta relação com o grau de confiança. Após a introdução deste teste, ele será utilizado no decorrer das seções para dar mais confiabilidade as conclusões apresentadas.

## 5 LÓGICA FUZZY EM SEGUROS

A primeira aplicação da Lógica Fuzzy em seguros aconteceu em 1982. Este artigo precursor buscou mensurar a imprecisão na subscrição, já que reconheceu que a subscrição era subjetiva. Propôs uma Máquina de Inferência Fuzzy para analisar a subscrição de uma companhia de seguros de vida (DEWIT, 1982). Primeiramente, foi solicitada a avaliação de 30 apólices de seguro de vida e as classificassem com base em vários atributos. As informações foram utilizadas para criar cinco termos linguísticos e respectivas funções de pertinência em função dos Conjuntos Fuzzy: Aspectos técnicos, saúde, profissão, comercial e outros,  $(\mu_t(x))$ ,  $(\mu_h(x))$ ,  $(\mu_p(x))$ ,  $(\mu_c(x))$  e  $(\mu_o(x))$ , respectivamente. A aplicação de DeWit combinou as funções de pertinências advindas dos Conjuntos Fuzzy segundo a expressão

$$W = \left( I(\mu_t)\mu_h\sqrt{\mu_p}\mu_h^2\sqrt{2\min(0,5;\mu_c)} \right)^{[1-\max(0;\mu_c-0,5)]}$$

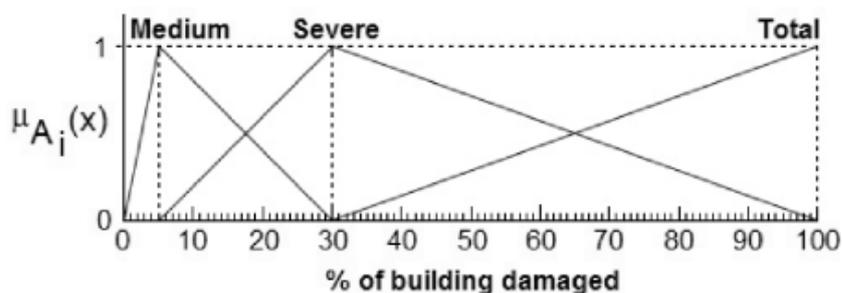
Os valores encontrados a partir desta equação foram aplicados na Máquina de Inferência Fuzzy, sugerindo as regras antecedentes e consequentes do sistema de decisão de Fuzzy para aceitação da apólice de risco.

- e  $0,0 \leq W < 0,1$  ; Então, recuse;
- Se  $0,1 \leq W < 0,3$  ; Então, tente melhorar as condições , se não for possível, recuse;
- Se  $0,3 \leq W < 0,7$  ; Então, tente melhorar as condições , se não for possível, aceite;
- Se  $0,7 \leq W < 1,0$  ; Então, aceite.

Em 1984, foi utilizada a Lógica Fuzzy para o reconhecimento de padrões na avaliação da atividade sísmica e previsão de danos, desenvolvendo modelos para estimar taxas de prêmios de seguros contra terremotos (BOISSONNADE, 1984). Este modelo possuía variáveis quantificáveis que podiam ser capturadas pelo Princípio de Extensão segundo modelos de probabilidade e variáveis categóricas foram melhor formuladas pelos Conjuntos Fuzzy que permitiram obter os valores de pertinências em função dos danos causados a um edifício pelo terremoto inferindo os Termos Linguísticos: médio, severo e total e a Máquina de Inferência Fuzzy propiciou a tomada de decisão quanto ao risco.

Após este artigo, as aplicações se expandiram consideravelmente em trabalhos envol-

Figura 9 - Fuzzy em seguros de terremotos



Fonte: (BOISSONNADE, 1984)

vendo classificação, passivos projetados, valores futuros e presentes, prêmios, alocações de ativos, fluxo de caixa e investimentos, utilizando os Conjuntos Fuzzy, Máquinas de inferências Fuzzy até Fuzzy Clustering (SHAPIRO, 2007).

A aplicação do Fuzzy Clustering, cidades foram agrupadas em delimitações geográficas segundo a frota de veículos em Massachusetts, permitindo a classificação das reivindicações de seguro de acordo com o nível de suspeita de fraude. (DERRIG, 1994).

O estudo realizado em Boston envolveu 350 cidades e categorias foram inferidas: recursos de responsabilidade por lesões corporais (BI), proteção de lesões pessoais (PIP), responsabilidade por danos materiais (PDL), colisão abrangente e outra categoria compreendendo as cinco coberturas individuais combinadas. A classificação resultou em algumas cidades pertencentes a mais de um cluster (classe de risco), resultado perfeitamente coerente com a modelagem Fuzzy. Resultados semelhantes foram encontrados, mas foram videnciados problema da classificação dos modelos convencionais

(VERRALL; YAKOUBOV, 1999) dados de custo total de sinistros associado a mais de 50.000 políticas de automóveis. Este trabalho aponta que, embora outros métodos pudessem ser usados, a flexibilidade da abordagem Fuzzy tornou-se mais adequada para estabelecer faixas etárias dos condutores segurados, concentrando a aplicação em coberturas de danos materiais a automóveis e lesões corporais. O risco relativo de cada grupo constante na última linha da Figura 11 e o requisito que as faixas de idades sequenciais têm associação assemelhadas, permitiu concluir que os grupos 4 e 5 provavelmente deveriam ser agregados, por ter riscos relativos próximos. Observou-se que o algoritmo poderia ser aplicado a variável referentes a características de seguro, de veículos para avaliar a classificação de risco. A modelagem se adapta melhor ao campo de seguros e detrimento aos modelos estatísticos frequentemente usados pelas seguradoras..

Figura 10 - Fuzzy em faixa etária de condutores

Age Groupings							
<b>Group</b>	1	2	3	4	5	6	7
<b>Risk Cluster</b>	1	2	3	4	3	5	6
<b>Age</b>	-25	26-27	28-31	32-47	48-51	52-68	69-
<b>Relative Risk</b>	406	136	115	90	100	72	61

Fonte: (SHAPIRO, 2007)

## 6 ANÁLISE EXPLORATÓRIA DOS DADOS

A amostra utilizada neste trabalho foi obtida através do site do Departamento Aplicado a Finanças e Estudos Atuariais da Universidade de Macquarie (MACQUARIE, 2008), constituí-se de informações relacionadas à apólices de seguro de veículo no período de 12 meses de observação, totalizando 67.803 apólices, Figura 11.

Figura 11 - Variáveis da amostra

VARIÁVEIS	EXPLICAÇÃO DOS VALORES	TIPOS DE VEÍCULOS	IMAGEM DOS VEÍCULOS
Valor do veículo	Valor do veículo em \$ 10.000	Caminhão	
Exposição	Intervalo de [0,1], onde 0,5 representa 6 meses de cobertura	Conversível	
Ocorrência de sinistro	0 = não houve sinistro 1 = houve sinistro (variável dummy)	Cupê: automóveis de capota fixa geralmente para dois ocupantes ou 2+2 e de duas portas.	
Número de sinistros	Número de vezes que o risco coberto ocorreu	Hardtop: automóvel com um telhado sólido feito de um material rígido.	
Montante de sinistro	Valor em \$, gasto pela seguradora, em cada apólice com todos os sinistros	Hatchback: hatch, veículo de passeio com o porta malas compacto.	
Tipo do veículo	Existem 13 tipos de veículos segurados, para melhor compreensão será explicado e disponibilizada uma imagem de exemplo.	Minibus: micro-ônibus e vans de grande porte.	
Faixa etária do veículo	Informada em cada apólice de 1 à 4, compreendendo 4 faixas.	Motorhome	
Gênero do condutor	Feminino ou Masculino	Ônibus	
Área	Informada pelas letras: A, B, C, D, E e F.	Panel van: vans com pouco lugares, furgão.	
Faixa etária do condutor	Possuem 6 faixas informadas apenas pelos números de 1 à 6.	Roadster: veículo de 2 lugares sem teto fixo.	
		Sedan: veículo de passeio que possui mais espaço interno e no porta malas comparado com o hatch.	
		Station wagon: chamado de Perua no Brasil.	
		Utilitários: veículos de tarefas, militares, limpadores de rua...	

Fonte: O autor, 2020.

O pré-processamento do tratamento dos dados é um ponto de partida essencial para a

modelagem, pois inclui tratamento prévio dos dados. (TUKEY, 1977) se referia a esta abordagem como a possibilidade de formulação de hipóteses dos experimentos.

Neste momento são identificados as variações discrepantes, investigam-se a interdependência entre variáveis, procura-se identificar tendências. Nesta primeira etapa que se estuda os fenômenos e sugere-se modelos. Este capítulo consiste em organizar os dados coletados em tabelas e gráficos para propiciar interpretação.

A amostra é composta pela ocorrência ou não de sinistros e as cinco variáveis quantitativas, sendo três contínuas, Valor do Veículo, Exposição, Montante de Sinistros e duas discretas, número e ocorrências de sinistros. As quatro variáveis qualitativas em escala nominal são: Gênero e Faixa Etária do Condutor, Faixa Etária Veículo por tipo e Área Geográfica. O Modelo proposto utilizou apenas as variáveis contínuas desta amostra por serem consideradas de maior impacto para precificação de seguro e permitem avaliar a capacidade do modelo proposto em reconhecer padrões para as apólices. O número limitado de variáveis torna importantíssimo que o modelo consiga diferenciar o padrão das unidades de observação, pois podem ser alocadas a um mesmo grupo, o que está em conformidade com a metodologia Fuzzy.

## 6.1 Análise de dados gerais

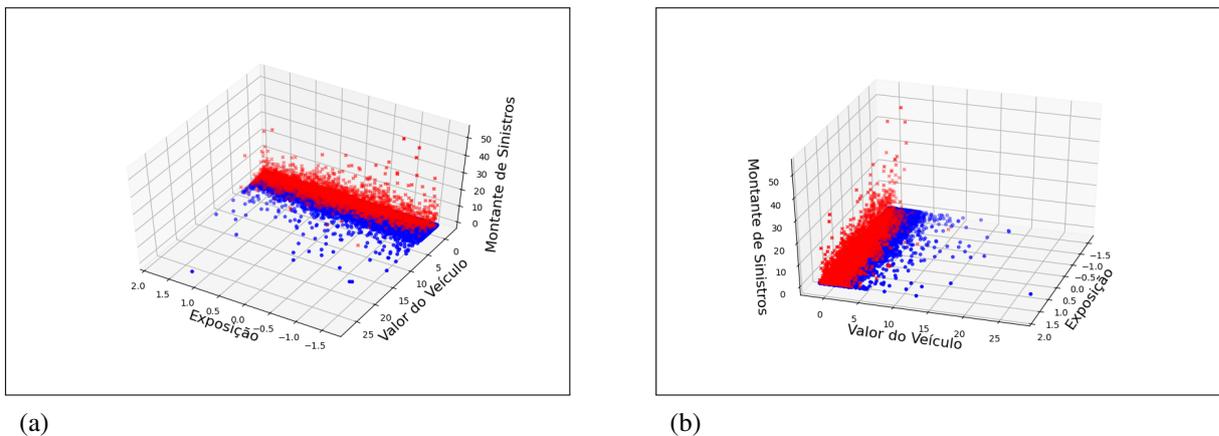
Como já mencionado, a carteira de seguros apresenta um total de 67.803 apólices onde 4.618 apresentaram pelo menos uma ocorrência de sinistro, 6,81% das apólices. Algumas obtiveram mais de um sinistro, totalizando 4.929 ocorrências e um montante de \$9.296.433,29, esta estatística é conhecida como risco médio ou esperança ( $E[S]$ ), onde  $[S]$  representa a variável aleatória valor total das indenizações ocorridas em uma carteira de seguros em um determinado período (FERREIRA, 2002).

O montante de sinistro, como mencionado anteriormente é o quanto a seguradora deverá arcar financeiramente com os segurados naquele período, em seguida a exposição, pois quanto maior a duração da vigência da apólice, mais tempo está exposto ao risco, conseqüentemente maiores chances de ocorrência de sinistros, e por fim a importância segurada, ou seja, o valor do veículo, uma vez que este é o argumento que está em risco. O gráfico a seguir apresenta todas as apólices da amostra em um gráfico sob a ótica destas três dimensões.

É importante ressaltar que as três variáveis, Exposição, Valor do Veículo e Montante de Sinistros estão padronizadas em unidades do desvio padrão, técnica para transformá-las na mesma escala de mensuração (LARSON; FARBER, 2008).

As apólices estão por toda a extensão do eixo Exposição, revelando que existem vigências de diversas mensurações. Naturalmente, as apólices indicadas pela cor azul permanecem na base do eixo Montante de Sinistros com o valor igual a zero, pois não houve ocorrências, enquanto as vermelhas crescem ao longo desse eixo. Outra observação é que o aumento do Valor do Veículo não determinou a ocorrência de sinistro. A avaliação da intensidade da associação

Figura 12 - Gráfico de Dispersão das apólices segundo Exposição, Valor do veículo e Montante de Sinistros padronizadas em unidades do desvio padrão segundo as perspectivas da Exposição e Valor do Veículo



Legenda: Azul: apólices sem ocorrência de sinistros. Vermelha: apólice onde ocorreu pelo menos 1 sinistro.  
Fonte: O autor, 2020.

monotônica, a que preservam a relação de ordem entre duas variáveis que tendem a mover-se na mesma direção relativa, mas não necessariamente a uma taxa constante, corresponde a correlação de Spearman (SPEARMAN, 1904), as a Coeficiente de Correlação de Pearson avalia relações lineares entre variáveis contínuas que se movem na mesma direção, a uma taxa constante, sendo que este coeficiente varia no intervalo  $[-1, 1]$ . A correlação entre Valor do Veículo e o Montante de Sinistro foi pequena, 0,02711. Na Tabela 5 está o resumo das estatísticas descritivas das variáveis objeto deste estudo.

Tabela 5 - Resumo das estatísticas gerais

	Montante de Sinistros	Exposição	Valor do Veículo
mean	\$137,11	0,468481	\$17.784,01
std	\$1.055,66	0,289907	\$12.046,78
min	\$0,00	0,002738	\$1.800,00
25%	\$0,00	0,219028	\$10.100,00
50%	\$0,00	0,446270	\$1.5000,00
75%	\$0,00	0,709103	\$21.500,00
max	\$5.5922,13	0,999316	\$34.5600,00

Fonte: O autor, 2020.

A Tabela 7 resume as estatísticas descritivas da amostra segundo as medidas de posição (média e mediana), de dispersão (desvio padrão e coeficiente de variação) e separatrizes (percentis 25, 50 e 75). A avaliação das estatísticas das 67.803 apólices da carteira permite perceber se o valor do veículo está acima da média, ou muito acima, ultrapassando o terceiro quartil, afere a variabilidade expressiva. É possível observar que o tempo mínimo e médio

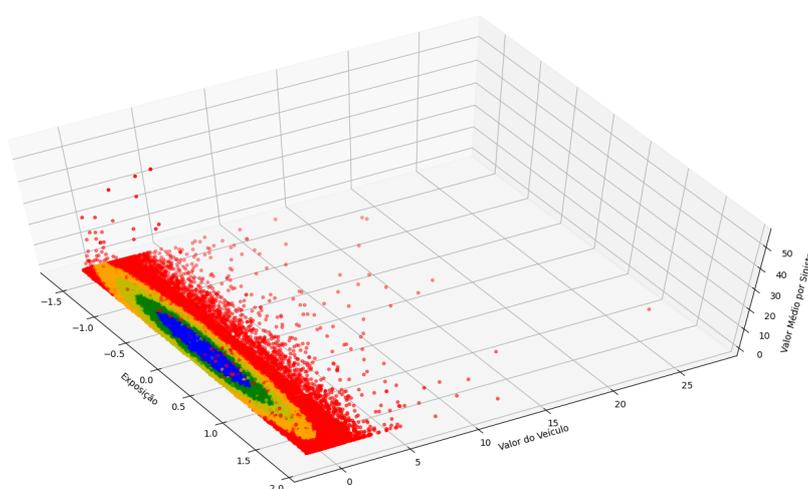
da Exposição ao risco foi de 1 dia e 6 meses, respectivamente. A análise estatística descritiva possibilita um olhar diferenciado do segurado e do analista em relação ao perfil da carteira em função das demais variáveis no sentido de decidir sobre a possibilidade de incluir mais variáveis de entradas ao modelo.

Na classificação de risco as variáveis foram padronizadas em unidades do desvio padrão para terem a mesma dimensionalidade. A Exposição é fator que aumenta as chances de ocorrência. A Importância Segurada afere o patrimônio envolvido quanto ao infortúnio e o Montante de Sinistros, os custos arcados pelos assumidores de risco quando migra para o fato consumado e aciona o compromisso da seguradora. Em seguida serão apresentados os resultados obtidos usando o Modelo Multivariado Logístico Fuzzy proposto .

## 7 RESULTADOS

O Modelo Multivariado Logístico Fuzzy foi implementado segundo a linguagem de programação Python (Anexo A) e indicou a classificação das apólices nas cinco faixas de risco: Baixíssimo, Baixo, Médio, Alto e Altíssimo, Figura 13.

Figura 13 - Distribuição das apólices por grupo de risco segundo Exposição, Valor do Veículo e Valor médio do sinistro

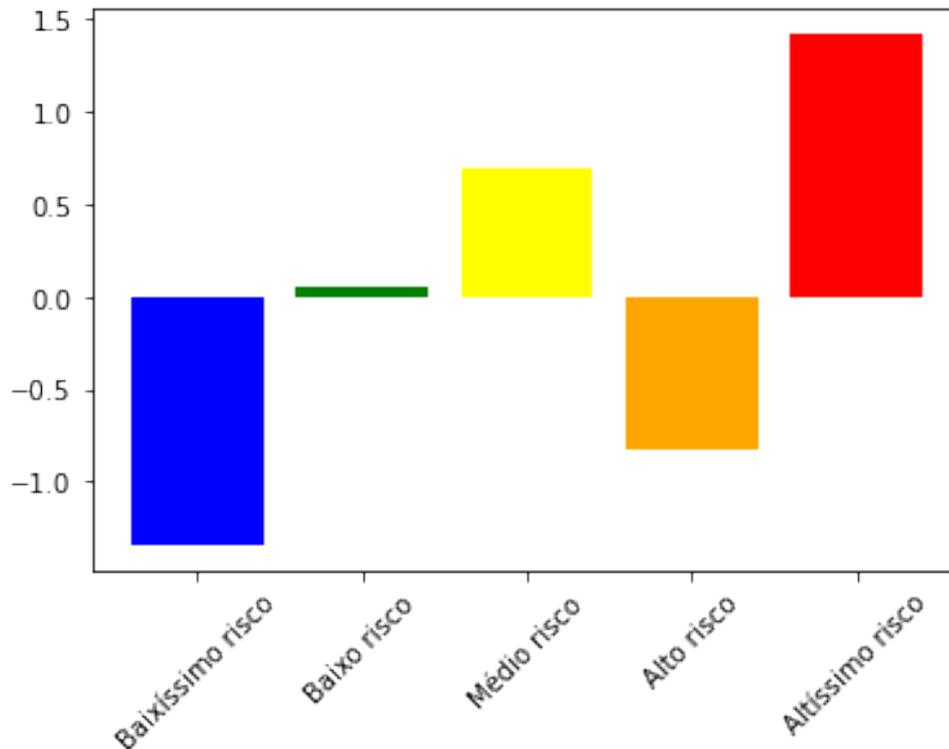


Fonte: O autor, 2020.

Esta classificação permitiu obter as distribuições de frequência das apólices para cada grupo de risco para as variáveis Exposição, Valor do Veículo e Valor Médio por Sinistro. O resultado mostrou a expressividade de apólice de Baixíssimo risco, tendo valores destas variáveis abaixo da média quando confrontado com o perfil dos demais grupos. A classificação revela também a pequena quantidade de apólices de Baixo e Alto risco, sendo que a frequência de apólices menores riscos conduz expectativa das Instituições na direção do gerenciamento de risco abaixo da média. A clareza na visualização destes grupos pode oportunizar uma política de bonificação para as Instituições. A Figura 15 apresenta a visualização das distribuições segundo os riscos padronizados em unidades do desvio padrão e observa-se que estão acima do risco médio são os de altíssimo risco, seguido de Médio Rico e com pouquíssima evidência, os de baixo risco. Os que se encontram abaixo do risco médio, estão o Baixíssimo risco e com incidência menor e de alto risco. Obviamente a expressividade de apólices de Altíssimo risco, não é um bom indicador e enfatiza a importância de uma política eficaz de reajuste para um bom gerenciamento dos riscos assumidos. Esta análise gráfica possibilita às Instituições optarem pela negação ou aceitação de novas apólices, pois é fácil interpretar a situação do risco já assumido

pela Instituição, pois avesso ao risco, a Instituição deve operar para manter a solvência, sendo capaz de arcar com suas responsabilidades junto ao segurado. Este resultado também permite a adoção de iniciativas de bonificação para apólices de Baixo Risco, possibilitando a cobrança justa pelo risco assumido e atrair novos segurados e de modo semelhante às apólices de Alto Risco, ocasionando aumentos de prêmio condizentes com o risco apresentado.

Figura 14 - Distribuição de Frequência dos riscos padronizados em unidades do desvio padrão segundo faixas de risco



Fonte: O autor, 2020.

Outro objetivo alcançado é a precificação dos prêmios dos seguros de cada apólice:

$$P = \mu_x * V * IS \quad (14)$$

Onde,  $P$  é o prêmio, valor pago pelo segurado;  $\mu_x$  é o valor da função de pertinência da apólice;  $V$  é o preço fixado para cada faixa de risco e  $IS$  é a importância segurada.

A equação proposta é um exemplo de precificação, diferentemente dos modelos que estão sendo aplicados, pois a abordagem Fuzzy cria uma diferenciação dentro das próprias faixas de risco, permitindo a mensuração de quanto o objeto pertence ao grupo.

Acresce que as apólices de mesma característica são precificadas igualmente. Na abordagem proposta consegue-se calcular as diferenças entre os perfis dos indivíduos que possuem a característica similar. Isso é de grande importância, uma vez que não tem sido mensurada

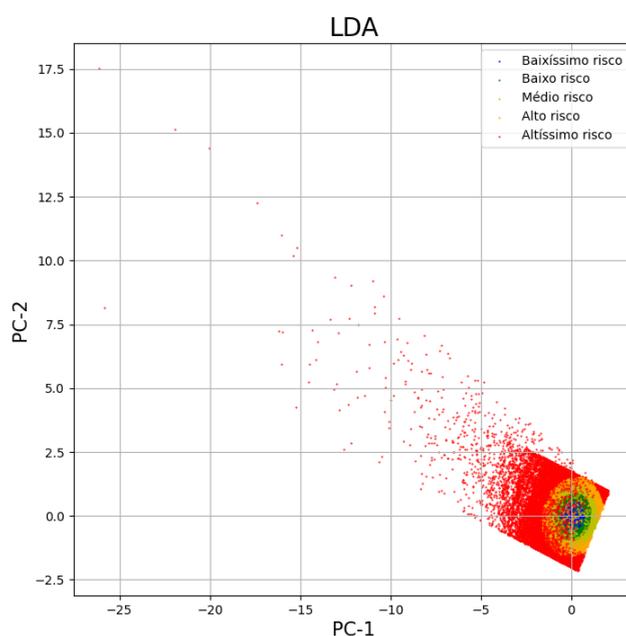
pelos modelos clássicos. A proposta desta dissertação proporciona uma precificação mais justa, seguindo uma das principais premissas da precificação: a cobrança baseada no verdadeiro risco assumido.

## 7.1 Validação do Modelo

Para a validação do modelo proposto, utilizou-se o Modelo de Análise Linear Discriminante (LDA) tendo com variável resposta a classificação em grupos de risco e as variáveis explicativas: onde cada componente principal é uma combinação linear de todas as variáveis originais

A Figura 15, a seguir, mostra esses componentes com as cores dos grupos classificados pelo Modelo Multivariado Logístico Fuzzy, tornando mais evidente a diferença entre os modelos.

Figura 15 - Análise Linear Discriminante para cinco agrupamentos em função das Componentes Principais, variáveis PC-1 e PC-2



Fonte: O autor, 2020.

Os modelos indicam diferentes distribuições de apólices nas respectivas faixas de risco, Tabela 6. O objetivo da validação do modelo é testar se as distribuições das classes nos modelos tem a mesma distribuição de frequência, ou seja, se os modelos convergem. Nesta comprovação foi aplicado o teste Qui-quadrado para verificar a aderência do modelo proposto Modelo

Multivariado Logístico Fuzzy com o Modelo de Análise Linear Discriminante, se os modelos convergem sob a Hipótese Nula: as distribuições são semelhantes. O teste foi realizado para o nível de significância  $\alpha = 0,05$ . Rejeita-se a Hipótese Nula se valor da estatística teste Qui-quadrado calculada for maior que o valor tabelado Qui-Quadrado para 4 graus de liberdade. O ponto crítico encontrado na tabela é de 9,49 e para os dados observados, foi encontrado o valor de 2,99. Dessa forma, com o valor menor que o ponto crítico, aceita-se a hipótese nula com  $\alpha = 0,05$  para 4 graus de liberdade, indicando a convergência entre os modelos, validando o Modelo proposto.

Tabela 6 - Distribuição das apólices entre os grupos de risco para Modelo Multivariado Logístico Fuzzy e o Modelo de Análise Linear Discriminante

<b>Modelo</b>	<b>Baixíssimo</b>	<b>Baixo</b>	<b>Médio</b>	<b>Alto</b>	<b>Altíssimo</b>
MLF	13529	13562	13577	13541	13594
LDA	13628	13626	13569	13549	13431

Fonte: O autor, 2020.

## CONSIDERAÇÕES FINAIS

Adotando o Modelo com algoritmo Fuzzy Multivariado pelo princípio de Extensão com a Métrica Logística foi possível reconhecer padrões em apólices de seguros, gerando uma classificação em cinco faixas de risco: Baixíssimo Risco, Baixo Risco, Médio Risco, Alto Risco e Altíssimo risco, como proposto nos objetivos. Além disso, foi possível viabilizar a análise do risco assumido de forma dinâmica, ágil e de fácil entendimento segundo as estatísticas intra-grupos, reconhecendo que as apólices de baixo risco podem oportunizar possíveis políticas de bonificação, enquanto a de alto risco, sofrer reajustes. Incentiva a adoção de políticas de subscrição e análise individual, mas também perceber o comportamento individualizado do risco assumido pelas instituições, o que muitas das vezes não é alcançada nos modelos estatísticos utilizados pelas seguradoras.

Embora possa considerar que se trata de uma amostra pequena, todas as análises ao longo deste trabalho permitiram concluir que a classificação individual do risco é uma forma mais eficaz de reconhecimento de padrões para as seguradoras.

Em resumo, o modelo proposto possibilita a análise e gerenciamento a partir de uma visualização clara e ágil dos riscos assumidos, proporcionando a rapidez na tomada de decisão e eficiência nas políticas de gestão e subscrição do risco. Destaca-se a abordagem proposta da precificação, destoando pela capacidade de interpretar as diferenças mesmo dentro da mesma faixa de risco. Uma técnica que pode ser considerada de grande valia por se tratar de uma ação essencial para o equilíbrio financeiro na precificação das apólices de seguro, reajuste e bonificação, pois contribui para a solvência da Instituição.

## REFERÊNCIAS

- ALFAKIH, Abdo Y. *Euclidean Distance Matrices and Their Applications in Rigidity Theory*. 1<sup>a</sup> ed.. ed. [S.l.]: Springer International Publishing, 2018.
- BARROS, Laécio Carvalho de; BASSANEZI, Rodney Carlos. *Tópicos de lógica fuzzy e biomatemática*. Volume 5. Campinas, SP: Comissão de Publicações – IMECC, 2006. 354 p.
- BCB, Banco Central do Brasil. *Plano Real*. 2020. Disponível em: <<https://www.bcb.gov.br/controleinflacao/planoreal>>. Acesso em: 20 ago. 2021.
- BERNARDI, Luiz Antonio. *Política e formação de preços: uma abordagem competitiva sistêmica e integrada*. 1<sup>a</sup>. ed. São Paulo: Atlas, 1996. 355 p.
- BOISSONNADE, A. C. *Earthquake damage and insurance risk*. Dissertação (Ph. D) — Stanford University, CA, 1984.
- CNT. *Release à imprensa: Brasil precisa de mais de R\$ 1,7 trilhão para solucionar problemas e modernizar infraestrutura de transporte, afirma CNT*. Brasília, DF: Confederação Nacional de Transportes, 2018. Disponível em: <<https://cnt.org.br/agencia-cnt/brasil-precisa-1-7-trilhao-investimentos-release-plano-cnt-transporte-logistica>>. Acesso em: 22 out. 2019.
- COORDENACÃO DE APOIO À GESTÃO ESTRATÉGICA (COGET), SUPERINTENDÊNCIA DE SEGUROS PRIVADOS (SUSEP). *7º Relatório de Análise e Acompanhamento dos Mercados Supervisionados*. Rio de Janeiro: SUSEP, 2019. 24 p. Relatório Técnico.
- CORIOLOANO, Marcio. *Sintonizado no Seguro com Marcio Coriolano e Mara Luquet*. 2018. Apresenta reproduções virtuais de pinturas brasileiras. Disponível em: <<https://cnseg.org.br/noticias/sintonizado-no-seguro-com-marcio-coriolano-e-mara-luquet.html>>.
- DENATRAN. *Departamento Nacional de Trânsito, Estatísticas DENATRAN*. Brasília, DF: Ministério da Infraestrutura, 2020. Disponível em: <[https://infraestrutura.gov.br/index.php?option=com\\_content&view=article&layout=edit&id=8551](https://infraestrutura.gov.br/index.php?option=com_content&view=article&layout=edit&id=8551)>. Acesso em: 26 mar. 2020.
- DERRIG, Richard A. Fuzzy techniques of pattern recognition in risk and claim classification. In: . [S.l.]: 4th AFIR International Colloquium, 1994. v. 1, p. 141–171.
- DEWIT, G. W. Underwriting and uncertainty. insurance: Mathematics and economics. In: . 4. ed. [S.l.: s.n.], 1982. v. 1, p. 277–285.
- FERREIRA, Paulo Pereira. *Modelos de precificação e ruína para seguros de curto prazo*. 1<sup>a</sup>. ed. Rio de Janeiro: FUNENSEG - Fundação Escola Nacional de Seguros, 2002. 224 p.
- FISHER, Ronald Aylmer. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, p. 179–188, 1936.
- FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. *Anuário Brasileiro de Segurança Pública*. São Paulo, 2018. Disponível em: <<http://www.forumseguranca.org.br/wp-content/uploads/2019/03/Anuario-Brasileiro-de-Seguran%C3%A7a-P%C3%BAblica-2018.pdf>>. Acesso em: 18 set. 2019.

GREENWOOD, Michael S. Nikulin Priscilla E. *A Guide to Chi-Squared Testing*. [S.l.]: Wiley, 1996. (Wiley Series in Probability and Statistics).

HOAGLIN, David C. John w tukey and data analyses. *Institute of Mathematical Statistics*, v. 18, n. 3, p. 311–318, 2003.

JONG, PIET DE; HELLER, GILLIAN Z. *Generalized Linear Models for Insurance Data*. 1<sup>a</sup>. ed. [S.l.]: Cambridge University Press, 2008. 208 p.

LANZILLOTI, Regina Serrão; LANZILLOTI, Haydée Serrão. Análise sensorial sob o enfoque da decisão fuzzy. In: NUTRIÇÃO, Revista de (Ed.). v. 12, n.2. Campinas: PUC Campinas, 1999. p. 145–157.

LARSON, Ron; FARBER, Betsy. *Estatística aplicada*. 2<sup>a</sup>. ed. [S.l.]: Pearson Prentice Hall, 2008. 476 p.

MACQUARIE. *Data Sets, Car*. Sidney, Australia: Department of Applied Finance and Actuarial Studies, 2008. Disponível em: <[http://www.businessandconomics.mq.edu.au/our\\_departments/Applied\\_Finance\\_and\\_Actuarial\\_Studies/research/books/GLMsforInsuranceData/data\\_sets](http://www.businessandconomics.mq.edu.au/our_departments/Applied_Finance_and_Actuarial_Studies/research/books/GLMsforInsuranceData/data_sets)>. Acesso em: 17 jul. 2019.

MAHALANOBIS, P. C. On the generalized distance in statistics. In: . [S.l.]: Proceedings of the National Institute of Science, 1936. v. 2, p. 49–55.

MAMDANI, E.H.; ASSILIAN, S. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, v. 7, 1975. 1–13.

MARQUES, João; BARRETO, Guilherme; CORTEZ, Paulo. Sistema de inferência fuzzy para interpretação da frequência cardíaca fetal em exames cardiotocográficos. 01 2006.

MCNEILL, F. Martin; THRO, Ellen. *Fuzzy logic: a practical approach*. [S.l.]: Morgan Kaufmann Pub, 1994.

MITAIM, Sanya; KOSKO, Bart. The shape of fuzzy sets in adaptive function approximation. *IEEE Transactions on Fuzzy Systems*, v. 9, n. 4, p. 1–20, 2001.

POLÍCIA RODOVIÁRIA FEDERAL. *Dados Abertos - Acidentes*. Brasília, DF: Ministério da Justiça e Segurança Pública, 2020. Disponível em: <<https://portal.prf.gov.br/dados-abertos-acidentes>>. Acesso em: 30 mar. 2020.

RAO, C. Radhakrishna. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society*, v. 10, n. 2, p. 159–203, 1948.

SHAPIRO, Arnold F. An overview of insurance uses of fuzzy logic. In: . [S.l.]: Springer Berlin Heidelberg, 2007. v. 2, p. 1–61.

SIEGEL, Sidney; CASTELLAN, Jr. N. John. *Estatística Não-Paramétrica para Ciências do Comportamento*. 2. ed. [S.l.]: Grupo A, 2006. 448 p.

SPEARMAN, Charles. The proof and measurement of association between two things. *The American Journal of Psychology*, p. 30, jan. 1904. Pp. 72-101.

SUPERINTENDÊNCIA DE SEGUROS PRIVADOS. *MINISTÉRIO DA FAZENDA, Superintendência de Seguros Privados, CIRCULAR SUSEP Nº 306*. Brasil, 17 de novembro de 2005.

SWISS RE INSTITUTE. *World insurance in 2017: solid, but mature life markets weigh on growth*. Suíça: Swiss Re Institute, 2018. 51 p. Relatório Técnico.

TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its applications to modeling and control. In: . v. 15, n.1. [S.l.]: IEEE transactions on systems, man, and cybernetics, 1985. p. 116–132.

TUKEY, John W. *Exploratory Data Analysis*. 1<sup>a</sup>. ed. [S.l.]: Addison Wesley, 1977.

VERRALL, R. J.; YAKOUBOV, Y. H. A fuzzy approach to grouping by policyholder age in general insurance. In: . [S.l.]: Journal of Actuarial Practice, 1999. v. 7, p. 181–203.

ZADEH, Lofti A. Fuzzy sets. *INFORMATION AND CONTROL*, n. N<sup>o</sup>.8, p. 338–353, nov. 1965.

\_\_\_\_\_. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE TRANSACTION ON SYSTEMS, MAN, AND CYBERNETICS*, n. SMC-3, N<sup>o</sup>.1, p. 28–44, jan. 1973.

ZIMMERMANN, H.J. *Fuzzy set theory - and its applications*. 4<sup>a</sup>. ed. New York: Springer Science+Business Media, LLC, 2001. 525 p.

## ANEXO A – Algoritmos

### A.1 Bibliotecas utilizadas

```

1 import matplotlib.pyplot as plt
2 import numpy as np
3 from numpy.linalg import inv
4 import pandas as pd
5 from scipy import stats
6 from scipy.spatial import distance
7 from scipy.stats import chisquare
8 from sklearn.discriminant_analysis import LinearDiscriminantAnalysis as
  LDA
9 from sklearn import metrics
10 from sklearn.model_selection import train_test_split
11 from sklearn.neighbors import KNeighborsClassifier
12 from sklearn.preprocessing import

```

### A.2 Normalização das variáveis

```

1 dados = pd.read_excel('data.xlsx') # Carrega os dados do excel para a
  variavel dados
2
3 # Transforma as respectivas colunas em vetores
4 IS = np.array(dados['Valor do Veiculo'])
5 vs = np.array(dados['Montante de Sinistros'])
6 exp = np.array(dados['Exposicao'])
7
8 # Normaliza os vetores
9 x = np.array(stats.zscore(np.array(dados['Montante de Sinistros'])))
10 y = np.array(stats.zscore(np.array(dados['Valor do Veiculo'])))
11 z = np.array(stats.zscore(np.array(dados['Exposicao'])))
12
13 df = pd.DataFrame([x,y,z]) # Organiza os vetores em uma tabela

```

### A.3 Distância de Mahalanobis

```

1 # Armazena a inversa da covariância dos dados normalizados para utilizar
  na funcao de distancia de Mahalanobis na variavel inv_cov
2 inv_cov = inv(df.T.cov())
3
4 #Funcao de distancia de Mahalanobis
5 def distancia(a, b, c, m, iv, x):
6     for i in range(len(dados['Exposicao'])):
7         x.append(distance.mahalanobis([a[i],b[i],c[i]],m, iv))

```

```

8     return x
9
10 a= []
11 m = [0, 0, 0]
12 distancias_m = distancia(x,y,z,m,inv_cov,a) # Amazena o resultado das
    distancias em distancias_m

```

#### A.4 Cálculo da Função de Pertinência

```

1 #Funcao que calcula a pertinencia utilizando o principio de extensao na
    Distancia de Mahalanobis
2 def pertinencia(x, distancia):
3     for i in range(len(dados[' rea '])):
4         x.append(1/(1+1/np.exp(distancia[i])))
5     return x
6
7 pert =[]
8 valores = pertinencia(pert, distancias_m)

```

#### A.5 Classificação dos grupos de Risco

```

1
2 def classificacao(valores): #Funcao para classificacao dos riscos em 5
    perfis
3     x =[]
4     for valor in valores:
5         if valor <= np.percentile(valores, 20): #Se menor igual a
            0.6732646403834693 primeiro percentil) Baix ssimo risco
6             x.append('Baix ssimo risco')
7         elif valor <= np.percentile(valores, 40):
8             x.append('Baixo risco') #Se menor igual a 0.7403367481024615
            (segundo percentil) Baixo risco
9         elif valor <= np.percentile(valores, 60):
10            x.append('M dio risco') #Se menor igual a
            0.7933307806000427 (terceiro percentil) M dio risco
11        elif valor <= np.percentile(valores, 80):
12            x.append('Alto risco') #Se menor igual a 0.8426913928474642
            (quarto percentil) Alto risco
13        else:
14            x.append('Alt ssimo risco') #Se maior igual a
            0.8426913928474642 (quinto percentil) Alt ssimo risco
15    return x

```

#### A.6 Análise Linear Discriminante

```

1 lda = LDA(n_components=2)
2 reduced_data = lda.fit(x,y).transform(x)

```

```

3 principalDf = pd.DataFrame(data = reduced_data, columns = ['PC-1', 'PC-2
  '])
4 finalDf = pd.concat([principalDf, df[['Classificacao']]], axis = 1)
5
6 df = finalDf
7 features = ['PC-1', 'PC-2']
8 X = df.loc[:, features].values
9 Y = df.loc[:, ['Classificacao']].values
10
11 X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size =
  0.3, random_state = 50000)
12 y_train=y_train.ravel()
13 y_test=y_test.ravel()
14
15 model = KNeighborsClassifier()
16 model.fit(X_train, y_train)
17 Yhat = model.predict(X_test)
18 print('*'*18, 'Accuracy', '*'*18, '\n')
19
20 for K in range(5):
21     K_value = K+1
22     neigh = KNeighborsClassifier(n_neighbors = K_value, weights='uniform',
  algorithm='auto')
23     neigh.fit(X_train, y_train)
24     y_pred = neigh.predict(X_test)
25     print ("Acuracia = ", metrics.accuracy_score(y_test,y_pred)*100, "% for
  K-Value:",K_value)

```

## A.7 Gráficos

```

1 # Plota grafico 3d com a classificacao das apolices em grupos de risco
2 color=[]
3 for i in range(len(pert)):
4     if pert[i] <= np.percentile(pert, 20):
5         color.append('b')
6     elif pert[i] <= np.percentile(pert, 40):
7         color.append('g')
8     elif pert[i] <= np.percentile(pert, 60):
9         color.append('y')
10    elif pert[i] <= np.percentile(pert, 80):
11        color.append('orange')
12    else:
13        color.append('r')
14
15 fig = plt.figure()
16 ax = fig.add_subplot(111, projection='3d')
17 x.scatter(z, y, x, s=10, c=color, marker='.')

```

```
18
19 ax.set_xlabel('Exposicao')
20 ax.set_ylabel('Valor do Veiculo')
21 ax.set_zlabel('Montante de Sinistro')
22
23 plt.show()
24
25 def grafico2(x,y,z):
26
27     m = list(range(len(distancias_m)))
28
29     col = []
30     t = []
31     for i in range(len(vms)):
32         if vms[i] > 0:
33             col.append('r')
34
35         else:
36             col.append('b')
37
38     plt.scatter(m, np.log(distancias_m), s=10, c=col, marker='o', alpha
39 =0.5)
40     plt.xlabel('Ordem dos valores')
41     plt.ylabel('Log da Distancia de Mahalanobis')
42
43     plt.show()
```