

UNIVERSIDADE CANDIDO MENDES – UCAM
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE PRODUÇÃO

Vitor Carneiro Escocard

UM ESTUDO SOBRE SEGURO DE AUTOMÓVEIS NA REGIÃO
SUDESTE DO BRASIL

CAMPOS DOS GOYTACAZES, RJ
Maio de 2017

UNIVERSIDADE CANDIDO MENDES – UCAM
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE PRODUÇÃO

Vitor Carneiro Escocard

UM ESTUDO SOBRE SEGURO DE AUTOMÓVEIS NA REGIÃO
SUDESTE DO BRASIL

Dissertação apresentada ao Programa de Mestrado em Engenharia de Produção da Universidade Candido Mendes – Campos/RJ, para obtenção do grau de MESTRE EM ENGENHARIA DE PRODUÇÃO.

Orientador: Prof. Ítalo de Oliveira Matias, DSc.

CAMPOS DOS GOYTACAZES, RJ
Maio de 2017

FICHA CATALOGRÁFICA

E74e Escocard, Vitor Carneiro.

Um estudo sobre seguro de automóveis na região Sudeste do Brasil. /. Vitor Carneiro Escocard – 2017.

92 f. il.

Orientador: Ítalo de Oliveira Matias

Dissertação apresentado ao Curso de Mestrado em Engenharia de Produção da Universidade Candido Mendes - Campos dos Goytacazes, RJ, 2017.

Bibliografia: f.83-92

1: Seguro de veículo. 2. Bibliometria. 3. Mineração de dados. 4. Acidente I. Universidade Candido Mendes – Campos. II. Título.

CDU – 368.03: 629.45

VITOR CARNEIRO ESCOCARD

**UM ESTUDO SOBRE SEGURO DE AUTOMÓVEIS NA REGIÃO
SUDESTE DO BRASIL**

Dissertação apresentada ao Programa de Mestrado em Engenharia de Produção da Universidade Candido Mendes – Campos/RJ, para obtenção do grau de MESTRE EM ENGENHARIA DE PRODUÇÃO.

Aprovada em: 12 de maio de 2017.

BANCA EXAMINADORA

Prof. Ítalo de Oliveira Matias, D.Sc. – Orientador
Universidade Candido Mendes

Prof. Aldo Shimoya, D.Sc.
Universidade Candido Mendes

Prof. Dalessandro Soares Vianna, D. Sc.
Universidade Federal Fluminense

CAMPOS DOS GOYTACAZES, RJ
2017

Dedico esse trabalho primeiramente a Deus, por ter me dado força de vontade para estar aqui e a minha família que tanto me apoia em minha carreira acadêmica.

Agradeço a Deus pelo suporte nos momentos de tormentos e por ser meu alicerce mais consistente em todo tempo, graças a permissão dele vim a esse mundo e graças a ele me tornei quem sou.

Agradeço a minha mãe por sempre estar ao meu lado me incentivando, desde os estudos mais intensos no ensino médio, passando pelo ingresso na Universidade Candido Mendes com 17 anos e agora buscando a minha formação como Mestre em Engenharia de produção com 24 anos.

Em especial agradeço meu irmão Fábio Carneiro Escocard que é Engenheiro de Produção e me auxiliou com as planilhas de excel e gráficos, por possuir um vasto conhecimento e já ter dado aula sobre isso, foram de grande valia seus ensinamentos.

Agradeço a meu Orientador, por de forma simples e clara sempre estar ao meu lado com intuito de aprimorar meu conhecimento e mais ainda pela sua disponibilidade de contato não só de forma pessoal, mas também se adequando aos meus horários com “plantões” por Skype e Whatsapp.

Agradeço a minha namorada por ter sido tão compreensiva durante todo esse tempo, onde por muitas vezes virou a noite ao meu lado me incentivando, ajudando nas pesquisas e me mantendo acordado, devido à pouca disponibilidade de tempo que possuo por conta da vida profissional, essas horas foram essenciais para conclusão deste.

Agradeço aos meus amigos que por várias vezes não pude oferecer a atenção que mereciam por estar com o tempo tomado e que mesmo assim sempre quando eu precisava espairar, estavam dispostos a sorrir comigo e me fazer ter certeza que todo esse esforço um dia valeria a pena, pois realizaria mais um sonho.

O caminho para a perfeição é fazer da dificuldade uma motivação.

Matheus Alves

RESUMO

UM ESTUDO SOBRE SEGURO DE AUTOMÓVEIS NA REGIÃO SUDESTE DO BRASIL.

A presente dissertação tem como objetivo fazer um estudo sobre o seguro de automóvel na Região Sudeste do Brasil, identificando qual faixa etária registrou mais sinistros nos últimos 10 anos, qual possuiu maior gasto médio por registro aberto, a que ofereceu maior prejuízo, a relação desses gastos para homem e mulher e para cada Estado da Região estudada. Na fase inicial do trabalho foi realizado um mapeamento dos artigos relacionados ao tema que se encontram indexados à base SCOPUS, a princípio foram localizados 521 artigos com as palavras chaves escolhidas que após aplicação de um método de filtragem, foi dada a origem de um ponto de partida composto por 50 artigos. Uma análise do banco de dados da DPRF (Departamento de Polícia Rodoviária Federal) referente aos dados de acidentes rodoviários dos últimos 9 anos (período disponível) permitiu verificar através da mineração de dados que a faixa etária de maior risco foi a de 26 a 35 anos, contradizendo os muitos autores que afirmam ser a faixa de 18 a 25 anos. Após levantamento desses dados, foi realizado uma cotação de seguro em 4 seguradoras diferentes modificando apenas o gênero do condutor (masculino e feminino), a idade do mesmo e o Estado onde se encontra o veículo (RJ; ES; MG; SP) que mostrou que em qualquer dessas situações a faixa etária que paga o maior valor de prêmio às seguradoras é a de 18 a 25 anos, o estudo mostra que esse método de cálculo observacional está ultrapassado.

PALAVRAS-CHAVE: Seguro de veículo. Bibliometria. Mineração de dados. Acidente.

ABSTRACT

AN ACADEMIC STUDY ON CAR INSURANCE IN THE SOUTHEAST OF BRAZIL

This paper aims at analyzing car insurance's characteristics in Brazil by identifying which age group claimed for sinisters the most within the last 10 years, which one caused the largest expenses, which caused the largest loss and relate these causes with gender and geographic location per state. Initially it was traced down all the articles related to the subject that are indexed to the SCOPUS basis and found 521 articles with the chosen key words. By sorting them out through data mining, it came down to a range of 50. A further analysis on the road accidents registered within the last 9 years extracted from the data base of the Federal Road Police Department will lead the reader to the similarities in the profile of the drivers from the 4 states of the region. Once similarities have been identified, this paperwork will bring on 4 different commercial quotations from 4 different insurance companies changing only the driver's gender, age and state where the vehicle is conducted, in order to point out the discrepancies between the quotations and the risks offered by each conductor.

KEYWORDS: Car insurance. Bibliometrics. Data mining. Accident.

LISTA DE FIGURA E GRÁFICOS

Figura 1.	Etapas do Processo KDD	31
Gráfico 1.	Quantidade de publicações por ano	25
Gráfico 2.	Quantidade de artigos publicados por afiliação	26
Gráfico 3.	Quantidade de artigos publicados por país	27
Gráfico 4.	Resultado da Análise para o Estado do Rio de Janeiro (RJ): de Pessoas do Gênero Masculino.	59
Gráfico 5.	Resultado da Análise para o Estado do Rio de Janeiro (RJ): de Pessoas do Gênero Feminino.	60
Gráfico 6.	Resultado da Análise para o Estado de São Paulo (SP): de Pessoas do Gênero Masculino.	61
Gráfico 7.	Resultado da Análise para o Estado de São Paulo (SP): de Pessoas do Gênero Feminino.	62
Gráfico 8.	Resultado da Análise para o Estado do Espírito Santo (ES) Pessoas do Gênero Masculino	63
Gráfico 9.	Resultado da Análise para o Estado do Espírito Santo (ES) Pessoas do Gênero Feminino	64
Gráfico 10.	Resultado da Análise para o Estado de Minas Gerais (MG) Pessoas do Gênero Masculino	65
Gráfico 11.	Resultado da Análise para o Estado de Minas Gerais (MG) Pessoas do Gênero feminino	66
Gráfico 12.	Análise com Gasto Médio (GM)	67
Gráfico 13.	Análise por Taxa de Acidente (TA)	69
Gráfico 14.	Análise do Prejuízo	71
Gráfico 15.	Análise por Amplitude	73

LISTA DE QUADRO E TABELAS

Quadro 1.	Diagrama de Pareto: Conceito	22
Tabela 1.	Distribuição da quantidade de publicações por tipo de documento	20
Tabela 2.	Distribuição dos títulos de periódicos por veículo de publicação	21
Tabela 3.	Distribuição dos artigos por autor	23
Tabela 4.	Dicionário de variáveis	43
Tabela 5.	Resultados obtidos no Estado de São Paulo (SP): gênero masculino.	50
Tabela 6.	Resultado obtido no Estado de São Paulo (SP): gênero feminino	52
Tabela 7.	Resultados obtidos no Estado do Rio de Janeiro (RJ): gênero masculino.	53
Tabela 8.	Resultados obtidos no Estado do Rio de Janeiro (RJ), gênero feminino.	53
Tabela 9.	Resultados obtidos no Estado de Minas Gerais (MG), gênero masculino.	55
Tabela 10.	Resultados obtidos no Estado de Minas Gerais (MG): gênero feminino	56
Tabela 11.	Resultados obtidos no Estado do Espírito Santo (ES): gênero masculino	57
Tabela 12.	Resultados obtidos no Estado do Espírito Santo (ES), gênero feminino.	58
Tabela 13.	Cotação para o gênero feminino no Estado do Rio de Janeiro	74
Tabela 14.	Cotação para o gênero feminino no Estado de Minas Gerais	75
Tabela 15.	Cotação para o gênero feminino no Estado de São Paulo	75
Tabela 16.	Cotação para o gênero feminino no Estado do Espírito Santo	76
Tabela 17.	Cotação para o gênero masculino no Estado do Rio de Janeiro	77

Tabela 18.	Cotação para o gênero masculino no Estado de Minas Gerais	77
Tabela 19.	Cotação para o gênero masculino no Estado de São Paulo	78
Tabela 19.	Cotação para o gênero masculino no Estado de Espírito Santo	78

LISTA DE ABREVIATURAS E SIGLAS

ANS - Agência Nacional de Saúde Suplementar

DM – Data Mining

DPRF – Departamento de Polícia Rodoviária Federal

GM – Gasto Médio (R\$)

PIB – Produto Interno Bruto

SUSEP – Superintendência de Seguros Privados

TA – Taxa de acidente (%)

SUMÁRIO

1.	INTRODUÇÃO	16
1.1.	CONTEXTO	16
1.2.	OBJETIVOS DA PESQUISA	18
1.2.1.	Objetivo Geral	18
1.2.2.	Objetivo Específico	18
1.3.	JUSTIFICATIVA	18
1.3.1.	Pesquisa na amostra	20
1.3.2.	Identificação dos periódicos com maior número de artigos publicados	20
1.3.3.	Identificação da evolução de publicações durante os anos	24
1.3.4.	Identificação das afiliações que mais publicaram artigos no assunto	26
1.3.5.	Identificação dos países que mais publicaram artigos no assunto	26
1.3.6.	Resultados obtidos: “identificação do núcleo de partida”	27
1.4.	QUESTÃO DA PESQUISA	30
1.5.	DELIMITAÇÕES DA PESQUISA	30
1.6.	ESTRUTURAÇÃO DO TRABALHO	30
2.	REVISÃO DE LITERATURA	32
2.1.	REFERENCIAL TEÓRICO	32
2.1.1.	O processo de descoberta de conhecimento em banco de dados (KDD).	32
2.1.2.	Clusterização de dados	34
2.1.3.	Algoritmo K-means	34
2.1.4.	WEKA	34

2.2.	ESTADO DA ARTE	35
2.2.1	Conclusão do Estado da Arte	40
3.	MATERIAIS E MÉTODOS	42
3.1.	UTILIZAÇÃO DO KDD PARA OS DADOS SELECIONADOS	42
3.1.1.	Seleção	42
3.1.2.	Pré-Processamento	44
3.1.3.	Transformação	44
3.1.4.	Mineração de Dados	44
3.2.	CÁLCULO DO PRÊMIO PAGO PARA CADA FAIXA ETÁRIA	45
3.3.	UTILIZAÇÃO DO KDD PARA OS DADOS SELECIONADOS	47
3.4.	CARACTERÍSTICAS QUE FORAM MANTIDAS EM TODOS OS PERFIS	48
4.	RESULTADOS	50
4.1.	RESULTADO DO SIMPLEKMEANS NO SOFTWARE WEKA	50
4.2.	RESULTADOS OBTIDOS A PARTIR DA BASE DE DADOS DA SUSEP	58
4.3.	RESULTADO DAS COTAÇÕES DE SEGURO	74
5.	CONSIDERAÇÕES FINAIS	79
5.1.	CONCLUSÕES	79
5.2.	SUGESTÃO DE TRABALHOS FUTUROS	82
6	REFERÊNCIAS BIBLIOGRÁFICAS	83

1. INTRODUÇÃO

1.1. CONTEXTO

O seguro de automóvel é um seguro personalizado que oferece aos segurados diversos serviços como atendimento 24 horas, indenização as vítimas em caso de acidente, reparo do veículo ou socorro mecânico, reboque do veículo e acidentes pessoais de passageiros. Esses são os 5 itens mais importantes oferecidos por todas as seguradoras conforme mostrado em Ribeiro e Grisi (1999), essas ofertas trazem aos condutores maior tranquilidade e planejamento financeiro na resolução de um problema.

O seguro é um contrato onde o segurado paga um valor chamado de “prêmio” à seguradora que irá assumir os custos financeiros no caso de um registro de sinistro (roubo, furto, batida, incêndio) além dos serviços que irá prestar como atendimento 24 horas, troca de pneu, reboque e outras particularidades de cada seguradora, quanto maior é o risco oferecido pelo segurado à seguradora maior será o valor do prêmio pago (MINUTOS SEGUROS, 2014).

O cálculo do seguro de um veículo no Brasil é baseado em diversos fatores como: endereço do segurado, local de pernoite, tipo de utilização do veículo, quilometragem média rodada por mês, se existe registro de seguro anterior ou é o primeiro, modelo do veículo utilizado, quantos veículos existem no local de pernoite, gênero do principal condutor e se existirá mais de um, se tem portão na garagem (automático, manual ou com vigia, tanto do local de pernoite quanto do trabalho caso utilize) e faixa etária do(s) condutor(es) (SEGUROFÁCIL, 2017).

O crescente lucro obtido pelas seguradoras provoca o crescimento de concorrência e surgimento de novas companhias de seguro, mas ainda assim, o número de segurados no Brasil é bem tímido perto da sua capacidade, aproximadamente 25% da frota de veículos possui seguro e esse número é capaz de gerar uma participação significativa de 6,2 % no PIB brasileiro (MENDONÇA, 2016).

Segundo o DPRF (2016) em 2015 foram registrados 269.052 acidentes, o que equivale em média a 22.421 acidentes por mês, 747 por dia e 31 registros por hora no Brasil, sendo eles de quaisquer naturezas que envolvam algum tipo de veículo. Essa quantidade de acidentes leva a insegurança de milhares de Brasileiros que utilizam veículo para trafegarem diariamente.

As seguradoras de automóveis buscam ligar diretamente o comportamento individual no trânsito à taxa de seguro paga pelos segurados (DIJKSTERHUIS et al. 2015).

Begg et al. (2015) afirmam que condutores jovens tem o maior risco de acidentes e o período de maior risco são os primeiros meses de direção sem supervisão de outras pessoas, os motivos deste risco elevado são os mais diversos, como inexperiência, falta de atenção, negligência, baixa percepção de risco, impulsividade, busca por emoção, dirigir após ingerir bebidas alcoólicas ou uso de drogas, entre outros.

Os acidentes de trânsito continuam sendo as principais causas de mortes em todo o mundo e a conduta do motorista é um fator contribuinte para 90 por cento dos acidentes na estrada, por isso, se faz muito importante a identificação do perfil dos condutores que contribuem para práticas inseguras no trânsito, buscando encontrar padrões que relacionem faixa etária, estado civil, quilometragem média percorrida diariamente e outros diversos fatores, para que esses sejam tratados e analisados de forma diferenciada (ELLISON et al. 2015).

1.2. OBJETIVOS DA PESQUISA

1.2.1. Objetivo Geral

Elaborar um estudo sobre os seguros de veículos na região Sudeste do Brasil, buscando mostrar que o método de predefinição de grupos de risco com base em dados observacionais utilizado pelas seguradoras no Brasil é impessoal e ultrapassado.

1.2.2. Objetivos Específicos

(I). Identificar o grupo que possui maior risco de se envolver em um acidente de trânsito com base em uma mineração de dados realizada nos registros do DPRF.

(II). Identificar em cada Estado da região Sudeste as faixas etárias que oferecem maior gasto médio por sinistro registrado, que mais se envolvem em acidente, as que oferecem maior prejuízo e as que possuem maior amplitude entre o menor e maior valor de prejuízo entre as faixas etárias.

(III). Avaliar a relação do cálculo de seguro elaborado por 4 seguradoras nos 4 Estados da Região Sudeste do Brasil, com os respectivos registros de sinistros disponibilizados pela SUSEP.

1.3. JUSTIFICATIVA

O setor de seguro de veículo possui um importante papel na economia brasileira, se observado a atual conjuntura econômica, o mercado de seguros vem obtendo grande avanço nos produtos comercializados o que reflete uma crescente participação no PIB do Brasil. Em 2015 o total de investimento do setor atingiu por volta de R\$ 803 bilhões, representando algo em torno de 14 % do PIB desse período (GERAL, 2016).

Segundo o Relatório de Sustentabilidade do Setor de Seguro (2016) em 2015, foi destacado que 63,3% do total de arrecadação do mercado de seguros do Brasil

está localizada na região Sudeste, o que incentivou a delimitar a pesquisa a esse local.

Devido a importância desse mercado para o Brasil, observando sua crescente participação no PIB, faz-se necessário observar se o método de análise de risco utilizada para o cálculo de apólices é eficiente, quanto mais justo for o prêmio pago pelo segurado maior valor percebido terá o consumidor final e o número de clientes tende a aumentar.

Para compor a justificativa do trabalho, uma pesquisa de Bibliometria foi realizada para definir a revisão bibliográfica do estudo, os resultados foram registrados com o intuito de se verificar os principais trabalhos que existem na área de seguro veicular, observar as opiniões dos autores e verificar que ainda assim existem poucos estudos nessa área publicados no Brasil, o que aumenta ainda mais o interesse em publicar tal dissertação, pois abrirá portas para novos estudos e trabalhos futuros. Serão destacados os principais autores que publicam nessa área, os países que já fizeram estudos similares, a evolução durante os anos de número de publicações, identificação dos anos que tiveram maior pico de publicação e os artigos de maior relevância julgados pela base SCOPUS.

A base de citações SCOPUS é uma fonte importante para pesquisadores, segundo a Elsevier (2011) a SCOPUS é a maior base de dados de citações e resumos da literatura peer-reviewed: revistas científicas, livros e anais de conferências. Ob-tendo-se uma visão global da produção de pesquisas, abrangendo todas as áreas do conhecimento. A base oferece ferramentas inteligentes para acompanhar, analisar e visualizar pesquisas.

Além de ser uma biblioteca virtual, possibilita a pesquisa através de filtros (que podem limitar sua pesquisa ou excluir dela) fazendo com que a busca ocorra de forma mais eficiente, possuindo também gráficos que quantificam os resultados ali obtidos podendo ser organizados como documentos publicados por ano, autor que mais publica, por país, por área de estudo, por tipo de documentos.

1.3.1. Pesquisa na amostra

A pesquisa foi realizada em 04/01/2017 através do portal de periódicos da capes, foi digitado o termo "risk driving" OR "car insurance" OR "seguro del automóvil" OR "vehicle insurance" OR "automobile insurance" devendo ser encontrado apenas em "Article Title, Abstract, Keywords", o filtro de ano não foi utilizado, a busca foi efetuada em todos os tipos de documentos encontrando-se assim 755 resultados, cuja distribuição por tipo de documentos encontrados está distribuída conforme Tabela 1 - Distribuição da quantidade de publicações por tipo de documento.

Forma de pesquisa: TITLE-ABS-KEY ("risk driving" OR "car insurance" OR "seguro del automóvil" OR "vehicle insurance" OR "automobile insurance").

Tabela 1 - Distribuição da quantidade de publicações por tipo de documento.

Tipo de documento	Quantidade de publicações
Article	521
Conference Paper	141
Review	42
Book Chapter	18
Article in Press	11
Book	8
Conference Review	5
Letter	4
Note	2
Short Survey	2
Business Article	1

Fonte: Adaptado pelo Autor ((2017) com dados a partir da Base Scopus)

Foram selecionados apenas os artigos para poder prosseguir com a análise bibliométrica, sendo assim, 521 ocorrências que geraram os resultados a seguir.

1.3.2. Identificação dos periódicos com maior número de artigos publicados.

De acordo com os parâmetros de busca, foram encontrados 521 registros de artigos na base de dados. A Tabela 2 mostra 216 artigos distribuídos conforme o veículo de publicação (uma representatividade de aproximadamente 41% do total), esses foram selecionados com o critério de 3 ou mais publicações.

Tabela 2 – Distribuição dos títulos de periódicos por veículo de publicação

TÍTULOS PERIÓDICOS	Nº PUBLICAÇÕES
ACCIDENT ANALYSIS AND PREVENTION	30
ASTIN BULLETIN	25
INSURANCE MATHEMATICS AND ECONOMICS	22
JOURNAL OF RISK AND INSURANCE	22
TRAFFIC INJURY PREVENTION	12
SCANDINAVIAN ACTUARIAL JOURNAL	9
TRANSPORTATION RESEARCH RECORD	8
GENEVA PAPERS ON RISK AND INSURANCE ISSUES AND PRACTICE	7
JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION	7
TRANSPORTATION RESEARCH PART F TRAFFIC PSYCHOLOGY AND BEHAVIOUR	6
JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY	5
EXPERT SYSTEMS WITH APPLICATIONS	4
JOURNAL OF SAFETY RESEARCH	4
JOURNAL OF TRAUMA INJURY INFECTION AND CRITICAL CARE	4
ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE	3
APPLIED ECONOMICS	3
EUROPEAN ACTUARIAL JOURNAL	3
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH	3
EVALUATION REVIEW	3
GENEVA PAPERS ON RISK AND INSURANCE THEORY	3
GENEVA RISK AND INSURANCE REVIEW	3
JOURNAL OF ADOLESCENT HEALTH	3
JOURNAL OF RISK AND UNCERTAINTY	3
JOURNAL OF SERVICES MARKETING	3
JOURNAL OF STUDIES ON ALCOHOL	3
NORTH AMERICAN ACTUARIAL JOURNAL	3
REVIEW OF ECONOMICS AND STATISTICS	3
REVISTA DE METODOS CUANTITATIVOS PARA LA ECONOMIA Y LA EMPRESA	3
RISK MANAGEMENT AND INSURANCE REVIEW	3
TRANSPORTATION RESEARCH PART A POLICY AND PRACTICE	3
XITONG GONGCHENG LILUN YU SHIJIAN SYSTEM ENGINEERING THEORY AND PRACTICE	3

Fonte: Adaptado pelo Autor ((2017) com dados a partir da Base Scopus)

Para agregar valor na interpretação dessa tabela, foi utilizado o diagrama de Pareto, que possibilita classificar por nível de importância os periódicos que devem ser monitorados futuramente com o intuito de verificar novos registros através dos principais veículos de publicações:

(I). Maior nível de atenção (Classe A): esses são os mais importantes e que devem ser monitorados, verificando se existirão novas publicações, pois devido a quantidade registrada tendem a ter ainda mais pesquisas. Em média 20% do total:

(A). Accident Analysis And Prevention; Astin Bulletin.

(II). Nível médio de atenção (Classe B): são de importância intermediária e frequentemente devem ser alvos de pesquisa buscando atualizações. Em média 30% do total:

(A) Insurance Mathematics And Economics; Journal Of Risk And Insurance; Traffic Injury Prevention; Scandinavian Actuarial Journal; Transportation Research Record.

(III). Os demais artigos (Classe C): possuem menor importância comparada aos listados anteriormente, pois registraram publicações de forma mais eventual. Em média 50% do total.

DIAGRAMA DE PARETO: Diagrama que permite selecionar itens de acordo com o seu nível de importância, determina dimensionar a ordem de prioridade dos fatores à que deve ser analisado ou monitorado (QUALIDADETOTAL, 2017)

Quadro 1. Diagrama de Pareto: conceito
Fonte: Elaborado pelo Autor (2017).

A Tabela 3 irá apresentar os autores com maior número de publicações sendo listados apenas os que publicaram 3 ou mais artigos. Foram observadas todas as novas publicações desses autores para verificar a possibilidade de estarem escrevendo novos estudos sobre o tema em questão.

Tabela 3 - Identificação dos autores com maior número de artigos publicados

Autores	Nº de publicações
Shope, J.T.	10
Tennyson, S.	8
Valdez, E.A.	8
Gómez-Déniz, E.	7
Li, C.S.	7
Liu, C.C.	7
Simons-Morton, B.G.	7
Ayuso, M.	6
Derrig, R.A.	6
Guillén, M.	6
Dedene, G.	5
Raghunathan, T.E.	5
Bingham, C.R.	4
Dionne, G.	4
Frees, E.W.	4
Shi, P.	4
Bermúdez, L.	3
Browne, M.J.	3
Chiappori, P.A.	3
Cummins, J.D.	3
Daroonch, A.H.	3
Hartos, J.L.	3
Karlis, D.	3
Leaf, W.A.	3
Patil, S.M.	3
Pinquet, J.	3
Spindler, M.	3
Suissa, S.	3
Viaene, S.	3
Weiss, M.A.	3

Fonte: Adaptado pelo Autor ((2017) com dados a partir da Base Scopus)

1.3.3. Identificação da evolução de publicações durante os anos

O Gráfico 1 mostrou o levantamento cronológico de publicações de artigos ao longo dos anos, foi possível verificar que os dois primeiros artigos foram escritos em 1927 e o terceiro só teve publicação 5 anos depois, mas só passou a ter recorrência de publicações em todos anos posteriores em 1968 e assim foi aumentando até atingir o pico em 2015 com 40 publicações.

Ciclos de produção mais acentuados:

- 1960 a 1962
- 1988 a 1989
- 1991 a 1993
- 1996 a 1998
- 2001 a 2003
- 2005 a 2016

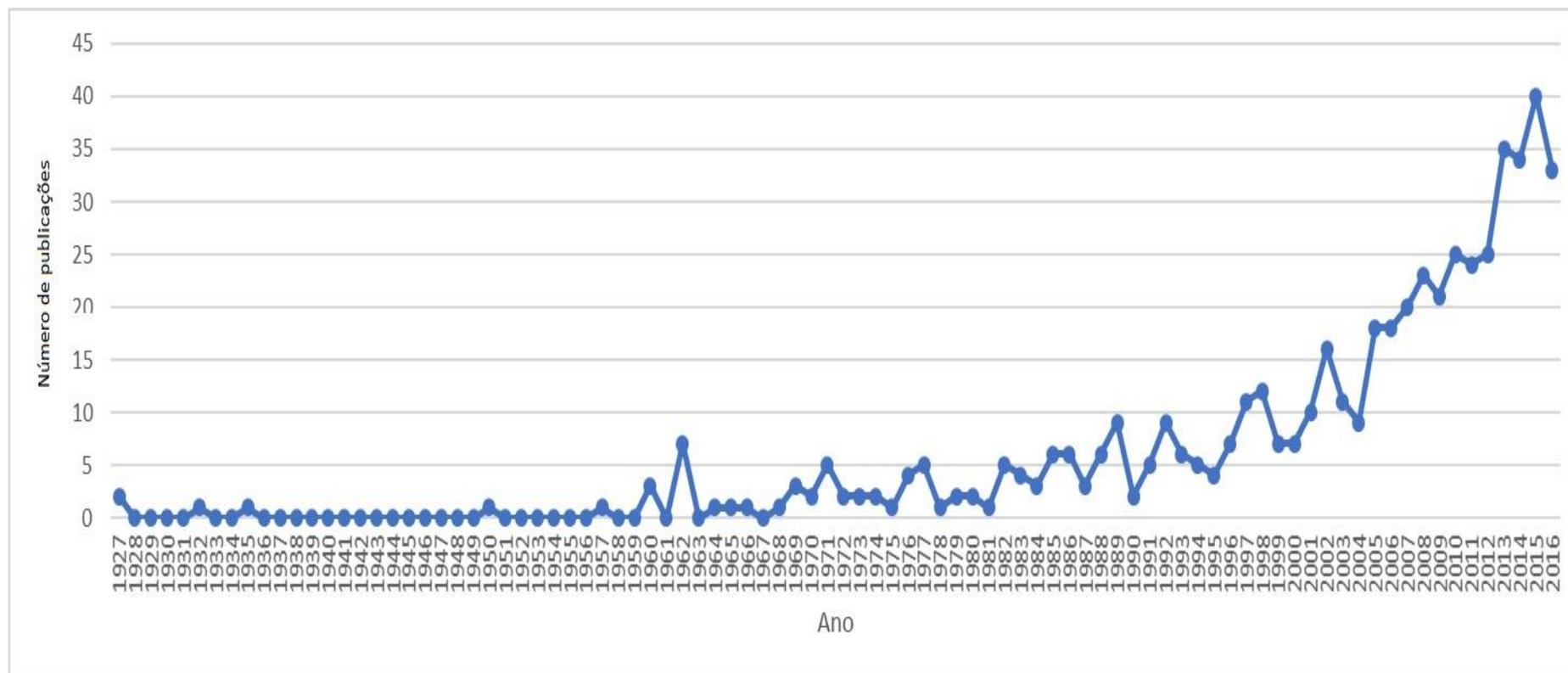


Gráfico 1 - Quantidade de publicações por ano

Fonte: Adaptado pelo Autor ((2017) com dados a partir da Base Scopus).

1.3.4. Identificação das afiliações que mais publicaram artigos no assunto

Nesse item foram selecionados apenas as 17 instituições que mais publicaram artigos no assunto, delimitados em 5 ou mais publicações conforme o Gráfico 2. A Universitat de Barcelona se destaca com 12 publicações, em seguida a Feng Chia University aparece com 10 e as demais com 8 ou menos.

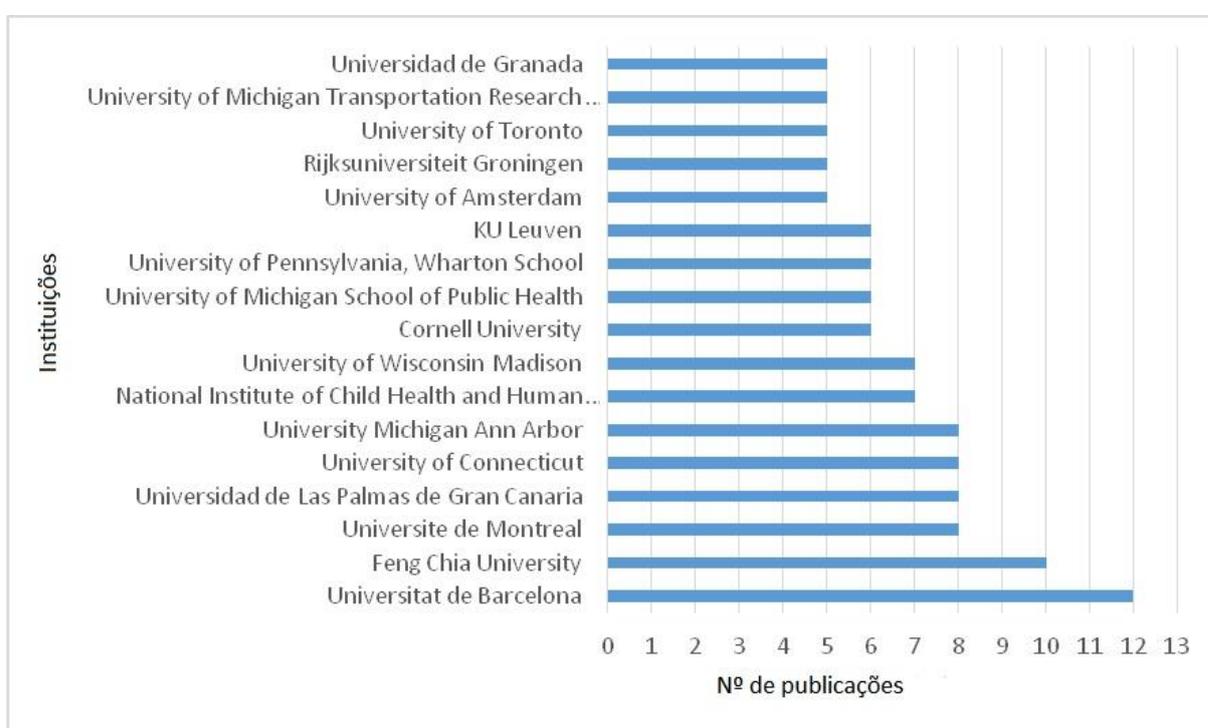


Gráfico 2 - Quantidade de artigos publicados por afiliação
 Fonte: Adaptado pelo Autor ((2017) com dados a partir da Base Scopus).

1.3.5. Identificação dos países que mais publicaram artigos no assunto

Os Estados Unidos lideram o ranking de publicações com 169 artigos, o Canadá aparece em segundo lugar registrando menos de 1/3 dessa quantidade. Foram selecionados apenas os que publicaram 10 ou mais artigos totalizando 14 países, para efeito de comparação, foi adicionado a quantidade de artigos publicados no Brasil que aparece em 26º colocado mostrando ser bem inferior aos demais, nota-se ainda que ao entrar em cada um desses 3 estudos publicados no

Brasil, percebe-se que em algum momento o seguro de carro foi citado mas não era especificamente uma pesquisa sobre o seguro o que comprova mais ainda a inferioridade desse quesito no país em relação aos demais. O gráfico 3 representa esses números.

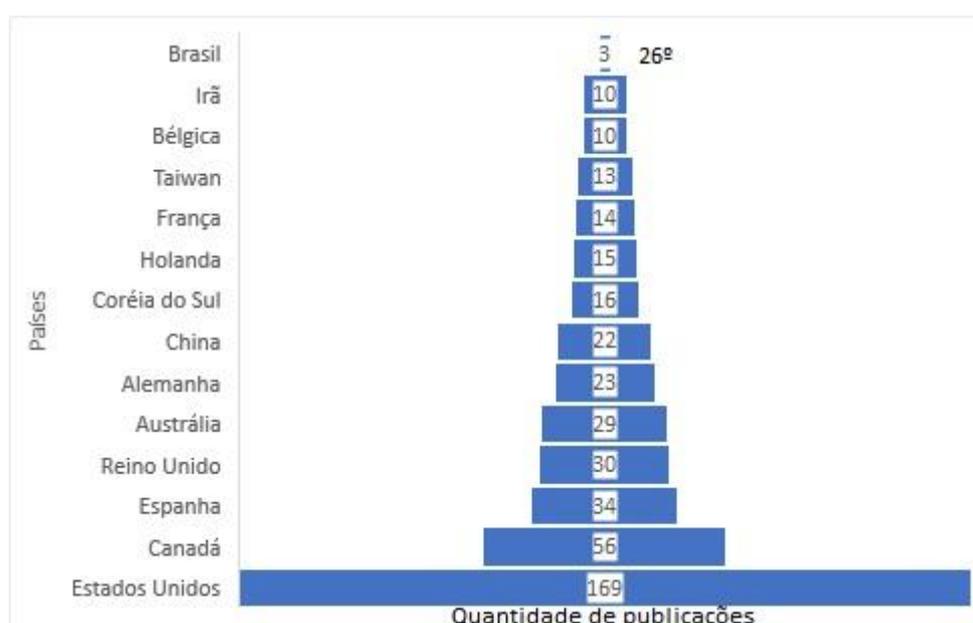


Gráfico 3 - Quantidade de artigos publicados por país
 Fonte: Adaptado pelo Autor ((2017) com dados a partir da Base Scopus)

1.3.6. Resultados obtidos: “identificação do núcleo de partida”

Para compor o núcleo de partida foram utilizados alguns critérios abaixo listados:

(1). 3 artigos mais antigos indexados na base – Busca-se identificar o objetivo dos estudos iniciais nessa área: (LOMAN, 1927, 1932; STELLWAGEN, 1927).

(2). 15 artigos mais recentes indexados na base - Foram analisados uma quantidade maior em comparação com os mais antigos para dar mais ênfase a esses, utilizando os mais antigos apenas como comparativos: (CANNON; CIPRIANI; BAZAR-ROSEN, 2016; GOLDEN et al., 2016; NANNI et al., 2016);(ARBIS; DIXIT; RASHIDI, 2016; BERNARD, 2016; DIJKSTERHUIS et al., 2016; GÓMEZ-DÉNIZ, 2016; HAUGEN; MOGER, 2016; HSU; CHOU; SHIU, 2016; LEE; WINSTON, 2016;

ROSS et al., 2016; SARABIA et al., 2016; SHI; ZHANG, 2016; STERN et al., 2016; WEIDNER; TRANSCHEL; WEIDNER, 2016).

(3). Seleção dos 15 artigos presentes na base com maior grau de relevância que é fornecida pela base SCOPUS: O artigo tido com maior grau de relevância não teve autor definido (“Automobile insurance”, 1965); (ADAMS, 1977; BAILEY; SIMON, 1960; DEVLIN, 1990; HARRINGTON; NIEHAUS, 1998; JOLOWICZ, 1971; JUNG, 1968; MARONEY; VICKORY, 1986; MARTIN, 1960; MUNDEN, 1962; PÉREZ-FRUCTUOSO; PÉREZ, 2010; SEGOVIA GONZÁLEZ; GUERRERO CASAS; HERRANZ PEINADO, 2007); já referenciadas (LOMAN, 1927, 1932; STELLWAGEN, 1927).

(4). Seleção de média de 10% dos artigos mais relevantes nos ciclos identificados na seção 1.3.3. buscando identificar quais artigos tiveram maior relevância em cada pico de publicação nos anos listados:

4.1. 1960 a 1962 – Artigos selecionados (2/10): (BAILEY; SIMON, 1960; MARTIN, 1960).

4.2. 1988 a 1989 – Artigos selecionados (2/15): (ATKINSON; NIGOL, 1989; CHAPPELL; NORMAN, 1989).

4.3. 1991 a 1993 – Artigos selecionados (2/20): (BEIRLANT et al., 1992; THOMAS; YOUNG, 1993).

4.4. 1996 a 1998 – Artigos selecionados (3/27): (HARRINGTON; NIEHAUS, 1998) que já foi citado anteriormente; (BLOCK, 1998; VATTER, 1997).

4.5. 2001 a 2003 – Artigos selecionados (4/35): (APTE et al., 2002; BLOWS et al., 2003; CUMMINS; PHILLIPS; WEISS, 2001; LI; LIU, 2003).

4.6. 2005 a 2016 – Artigos selecionados (10/312): (BOJ et al., 2007; BURLACU, 2012; LEDO, 2011; MANDIC, 2013; PÉREZ-FRUCTUOSO; PÉREZ, 2010; SEGOVIA GONZÁLEZ; CONTRERAS RUBIO; MAR MOLINERO, 2009; SEGOVIA GONZÁLEZ; GUERRERO CASAS; HERRANZ PEINADO, 2007; YU; HAO, 2008; YUZVOVICH et al., 2015; ZHAO; ZHOU, 2014)

O presente estudo bibliométrico atingiu o seu objetivo de identificar um núcleo de partida para novos estudos no tema de seguro de veículo, a princípio foram encontrados 571 artigos que obtinham uma das palavras chaves escolhidas como definição da amostra, isso mostra uma grande quantidade de estudos na área, o que por um lado traz muita informação que agrega qualidade a pesquisa, por outro torna difícil a análise de tamanha base de dados que apresentará um serviço mais maçante de busca e menos produtivo.

Ressalta-se que não foi encontrado nenhum artigo durante toda a pesquisa na base SCOPUS que fosse semelhante a esse, demonstrando o valor do cálculo realizado pelas seguradoras e comparando as faixas etárias entre si com os respectivos riscos oferecidos por cada uma delas.

O método proposto por Costa (2010) gerou 50 artigos que sugere como ponto de partida para realização desse estudo, os métodos de escolhas foram bem explicados e divididos com o intuito de se mostrar o porquê desses trabalhos terem sido selecionados.

O primeiro estudo foi publicado em 1927, o segundo em 1932 e foi se mantendo bem baixa a quantidade de publicações até 1968 quando começa a ter pelo menos uma publicação por ano ininterruptamente, mostrando o aumento da importância que esse mercado tem para os Países.

Em 2015 registrou o maior número de publicações de todos os anos, nesse ano o Brasil enfrentou umas das mais fortes crises da história do país, não só econômica como também política e quando se observa os principais indicadores macroeconômicos desse período, verifica-se uma predominância de índices negativos, incluindo a variação do próprio PIB, a desvalorização da moeda real em relação ao dólar, aumento da taxa de desemprego, inflação em ascensão e ainda assim o mercado de Seguro registrou segundo informações da Superintendência de Seguros Privados SUSEP e da ANS uma alta nominal de 11,6% do faturamento total em relação a 2014 (CNSEG, 2016).

O maior ganho que se tem com esses dados de partida é que eles além de oferecer as informações primordiais e darem um norte para a pesquisa, é deles que

serão extraídas outras referências bibliográficas para maior enriquecimento da pesquisa.

1.4. QUESTÃO DA PESQUISA

As questões centrais da pesquisa são: Qual a faixa etária que apresenta maiores riscos as seguradoras? O quão mais caro elas pagam na hora de contratar o seguro? O método utilizado no cálculo atual dos seguros no Brasil é eficiente?

1.5. DELIMITAÇÕES DA PESQUISA

A pesquisa foi realizada com os dados do SUSEP dos últimos 10 anos de todas as seguradoras de automóveis do Brasil e na opção de filtragem foram escolhidos apenas os dados do Sudeste do Brasil conforme foi o foco da pesquisa, porém, esses dados foram comparados ao cálculo online realizado por apenas 4 seguradoras, o que delimitou os resultados obtidos e uma quantidade significativa de dados do DPRF (cerca de 33%) foi excluída por terem sido divulgados como "não identificados".

1.6. ESTRUTURAÇÃO DO TRABALHO

O trabalho está estruturado em 5 capítulos:

(I). Capítulo 1: Introdução: Apresenta a importância do mercado de seguros para o Brasil, abordando como são realizados os cálculos de seguros, os objetivos da pesquisa, a estrutura do trabalho, a motivação do trabalho através de uma pesquisa bibliográfica para mapear estudos científicos sobre o tema, a questão do trabalho e fecha o capítulo com as delimitações encontradas e estruturação do trabalho.

(II). Capítulo 2: Revisão de Literatura: Serão abordadas literaturas que discorrem sobre os métodos utilizados na busca do conhecimento, as etapas de cada processo e os softwares utilizados. Também serão discriminados os resumos dos principais

artigos selecionados pelo método da pesquisa bibliométrica e por fim será realizada uma conclusão do capítulo.

(III). Capítulo 3: Materiais e Métodos: Trata-se da aplicação dos métodos descritos no capítulo 2, onde as etapas do KDD foram implantadas na base de dados da DPRF; mostra-se como foi efetuada a pesquisa na base de dados da SUSEP e como foi realizado o filtro da mesma, por fim demonstra como foi efetuado o cálculo dos seguros e os critérios utilizados.

(IV). Capítulo 4: Resultados e Discussões: Foram expostos os resultados obtidos na descoberta do conhecimento proporcionado pelo algoritmo SimpleKmeans no software WEKA; os gráficos que se originaram do banco de dados da SUSEP proporcionaram definir uma ordem de risco decrescente em relação a faixa etária e esses resultados foram comparados com os valores das cotações realizadas em busca de uma convergência de informações.

(V). Capítulo 5: Considerações finais: Nesse capítulo são apresentadas a resposta da questão da pesquisa, a conclusão, as considerações finais, as limitações e sugestões de trabalhos futuros.

2. REVISÃO DE LITERATURA

2.1. REFERENCIAL TEÓRICO

2.1.1. O processo de descoberta de conhecimento em banco de dados (KDD)

Na era da informação, o conhecimento vem se tornando um recurso de grande importância para as organizações, proporcionando uma vantagem competitiva frente às demais companhias. Muitas organizações têm coletado e armazenado uma grande quantidade de dados, porém nem todas as empresas utilizam e descobrem o conhecimento contido em todo esse banco de dados. Administrar o conhecimento é um grande desafio e o processo de KDD (Knowledge Discovery in Databases) vem auxiliando a descoberta de conhecimento nos banco de dados existentes (SILWATTANANUSARN, 2012).

Segundo Fayyad et al. (1996) o processo de KDD é iterativo e interativo, envolvendo algumas etapas e decisões tomadas pelo usuário.

Baseado na Figura 1 o processo de KDD consiste na seguinte sequência de métodos iterativos:

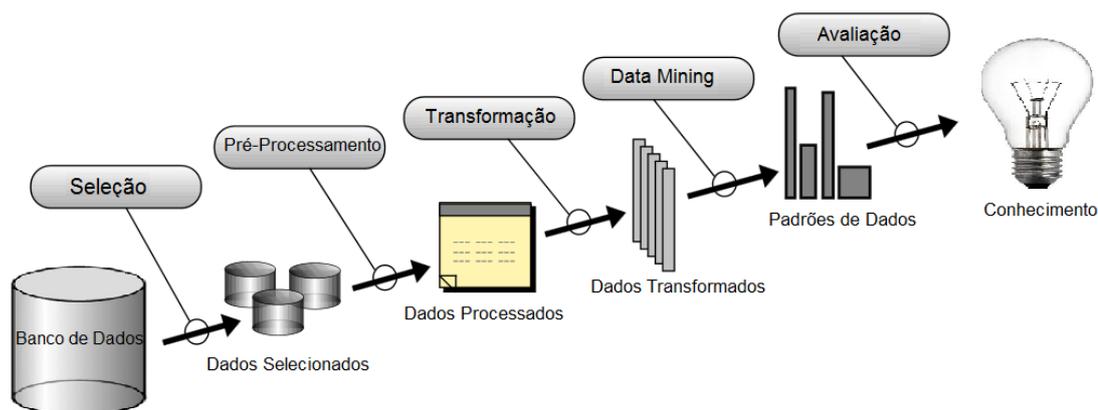


Figura 1 - Etapas do processo de KDD

Fonte: adaptado de Fayaad et al. (1996)

(1). Seleção: Selecionar dados relevantes no qual será feita a descoberta do conhecimento, esse processo inicial é importante, pois nele serão selecionados os dados que conterão todas as possíveis variáveis, o que irá impactar diretamente no resultado final, portanto aconselha-se ser feito por um especialista no assunto;

2. Pré-Processamento: Os dados passam por uma limpeza, eliminação de dados incompletos, dados redundantes e dados que não agregarão valor a pesquisa, essa fase é importante porque influenciará diretamente na qualidade dos resultados apresentados pelo algoritmo de mineração;

3. Transformação: Os dados são armazenados de acordo para facilitar a aplicação do algoritmo de mineração, devem ser transformados em formato compatível com o algoritmo que será utilizado para a próxima etapa.

4. Data Mining: Essa etapa se torna a de maior destaque entre as demais, o que não diminui a importância de cada uma, isso ocorre, pois nela que irá ocorrer a interpretação dos padrões e geração de conhecimento de dados, o DM é utilizada em grandes bases de dados e geralmente de maior complexidade.

5. Avaliação: Os resultados obtidos são avaliados e interpretados pelo usuário, essa etapa é onde ocorre o retorno as demais caso o resultado não tenha sido satisfatório.

2.1.2. Clusterização de dados

O processo de clusterização é uma tarefa descritiva comum onde se procura identificar um conjunto finito de grupos ou categorias para descrever os dados.

As categorias consistem em uma representação mais rica, como categorias hierárquicas ou sobrepostas (FAYYAD et al. 1996).

Segundo Jain (2008) a clusterização de dados pode ser utilizada para três principais propósitos:

(A). Estrutura principal: identificar características mais relevantes, para obter conhecimento sobre os dados, detectar anomalias e gerar hipóteses;

(B). Classificação Natural: identificar o grau de semelhança entre as formas ou organismos;

(C). Impressão: Método para organização dos dados e resume-os através de clusters.

2.1.3. Algoritmo K-means

O K-means é uma técnica de análise de cluster que possui o objetivo de agrupar algoritmos após a classificação de objetos com base em K números de clusters. O K-means é considerado como um método não supervisionado, ou seja, não há necessidade de nenhuma supervisão humana, a classificação ocorre de forma automática. O método Simple K-means é um dos algoritmos de clusterização mais utilizados, e se tornou popular pela sua velocidade e praticidade e será utilizado neste trabalho para descobrir grupos de perfis de condutores que são mais relevantes nos acidentes de trânsito.

2.1.4. WEKA

É um software de conjunto de algoritmos de aprendizado de máquina que é utilizado para fins de mineração de dados. O weka contém ferramentas que são

utilizadas no pré-processamento como classificação, regressão, clustering, regras de associação e visualização. Permite a execução dos algoritmos de data mining de forma interativa. Os dados devem estar no formato ARFF que é o padrão de estruturas de bases manipuladas pelo Weka (WAIKATO, 2017)..

2.2. ESTADO DA ARTE

Os estudos que mais se destacaram foram de Maroney e Vickory (1986), Golden et al. (2016) e Lee e Winston (2016) onde juntos apontaram os principais fatores que aqui foram abordados, logo, serão primeiramente detalhados antes dos demais que ofereceram algum tipo de contribuição porém em escala menor.

Maroney e Vickory (1986) publicaram um artigo que possuía o objetivo de verificar a legalidade da utilização de grupos pré-classificatórios utilizados pelas seguradoras para oferecer diferentes valores de prêmios para um mesmo seguro, se diferenciando apenas pelo gênero, idade, escolaridade ou estado civil. As seguradoras sustentam essa classificação informando que grupos mais jovens possuem maiores taxas de acidente que os mais velhos e que as mulheres se envolvem menos acidentes que os homens, que casados se arriscam menos que os solteiros e que quanto maior o grau de escolaridade menor o risco.

O estudo conclui que as seguradoras utilizam esses grupos por comodidade de classificação e pelos supostos grupos de alto risco subsidiarem os de menor risco, o sistema é desleal, há evidências de que as variáveis idade, gênero, grau de escolaridade e estado civil não refletem efeito de perda superior, funcionam apenas como substitutos das causas verdadeiras. Em vários estados já são proibidos ou vem tentando proibir algumas ou todas dessas variáveis como classificatórias de grupos, pelos registros estudados as tentativas isoladas na justiça de inibir essas práticas vêm obtendo resultados negativos, a assertividade aconteceria com uma intervenção do estado.

Golden et al. (2016) afirmam que embora as variáveis de subscrição pessoal tenham surgido no passado (idade, gênero, estado civil) para classificação de risco do segurado, com o avanço da tecnologia computacional tornou-se possível o uso

de variáveis menos observacionais. Utilizando de análises comportamentais referentes a finanças pessoais e histórico de crédito é possível se obter uma nova forma de análise de riscos, mas tiveram que esperar até que os dados de crédito estivessem imediatamente acessíveis e a velocidade computacional tenha sido suficiente para que sejam implementados em tempo real. Buscar compreender o estilo de vida de uma pessoa e as suas escolhas tomadas podem ser a chave para o avanço de descobertas de ligações entre as variáveis diversas (comportamentais e genéticas) e os riscos que cada pessoa oferece às seguradoras, sendo assim o estudo propõe evidências favoráveis sobre o uso de pontuações de crédito para previsões de perdas de seguros com justificativas psicossociais e explicações bioquímicas.

Lee e Winston (2016) propuseram um estudo com o objetivo de desenvolver e validar um simulador para induzir estados emocionais negativos em condutores jovens e verificar se existe uma correlação positiva entre a magnitude de um estado emocional ao conduzir um veículo e o número de comportamentos inseguros no trânsito. Concluiu então que o simulador era válido e que os jovens recém-licenciados não lidam bem com a presença de trânsito lento ou de freadas inesperadas dos carros que estão à frente, esses resultados sugerem que estudos nessa área devem ser aprimorados com o intuito de oferecer ajuda aos mais jovens condutores a aprenderem a controlar melhor suas emoções e lidar com situações de estresse no trânsito.

Os demais artigos que foram selecionados para compor o referencial teórico são aqui resumidos:

O primeiro estudo encontrado na base foi uma pesquisa realizada por (LOMAN, 1927), em que aborda que já em 1926 existiam 22.342.457 automóveis registrados nos Estados Unidos e que esse número representava um aumento de 11,2% em relação a 1925, o crescimento do setor trouxe junto consigo um aumento de mortes e acidentes de trânsito. No estudo ele propõe analisar a obrigatoriedade do seguro na cidade de Massachusetts, demonstra as vantagens e desvantagens desse modelo e conclui que grande parte da discussão foi realizada de forma dedutível, mas que é suficiente para se verificar que é uma questão importante que

deve ser estudada mais a fundo pois se trata de uma resolução de problemas sociais.

(LOMAN, 1932) realizou um novo estudo que demonstrava que a legislação adotada por Massachusetts em que o seguro de responsabilidade civil deveria ser obrigatório não aumentou o número de acidentes conforme alguns críticos previam, ele fez um alerta para a suspensão do plano de mérito que o Estado havia realizado em 1930, plano esse que previa uma classificação dos condutores e essa classificação aumentava ou reduzia o valor de prêmio pago. Por ser possivelmente uma discriminação foi suspenso, porém o autor afirma ser um bom sistema e que ele deveria ser aprimorado com análise de perfil por idade e registros de acidentes, mas apenas sugeriu no final não demonstrando nenhum dado.

(BAILEY; SIMON, 1960) propuseram um estudo que compara os métodos de classificação de risco separando-os em rating de mérito (método canadense) e plano de classe (método americano), um baseia-se na experiência de carros previamente classificados de acordo com os registros existentes, conhecido popularmente lá como “plano de avaliação de mérito” o outro analisará o perfil do condutor: gênero, idade, ocupação, uso do veículo, local onde se encontra. Concluiu que o rating de mérito é tão eficaz quanto o plano de classe, porém os dois deixam a desejar por deixarem de lado a avaliação da quilometragem percorrida pelo condutor e que essa variável é uma causa de variação entre os riscos individuais.

No estudo realizado por Martin (1960), já se notava uma tendência a discriminação aos jovens condutores menores de 25 anos ao se constatar que não era permitido ofertar desconto aos condutores que possuíam em sua família algum condutor com menos de 25 anos. A pesquisa realizada sobre o sistema canadense de classificação para o seguro de automóvel afirmou que após a reimplantação do sistema de classificação em 1952 que havia sido suspenso em 1930, houve uma redução dos números de registros de sinistros e concluiu afirmando que o regime permitiu a indústria dar reconhecimento valioso aos segurados cujos registros o justificam e que esse se tornou o permanente sistema de seguro no Canadá.

Munden (1962) propôs um estudo que elaborou algumas análises sobre o seguro de automóvel, buscando relacionar a quantidade de sinistro registrada com

os perfis de condutores, ele reconhece que as variáveis e conclusões desse estudo são controversas e dedutíveis. Foram analisados os registros do período de outubro de 1954 a dezembro de 1955 e constatou que a quantidade de sinistro foi ligeiramente menor para as mulheres do que para os homens, no primeiro ano de seguro há um maior índice de registro independentemente da idade do condutor, segurados que registravam sinistros no primeiro ano de apólice tendenciosamente voltaram a registrar alguns anos depois, a faixa etária tida como mais segura foi de 40 a 50 anos e fecha o estudo afirmando que o período estudado pode ter sido influenciado por alguns fatores da Segunda Guerra Mundial.

Devlin (1990) propôs um estudo na região de Quebec – Canadá, onde tinha o objetivo de verificar qual sistema é mais eficiente no ponto de vista econômico, o sistema sem culpa (seguradora arcará com o prejuízo se o condutor provar não ter tido culpa) adotado pela cidade ou o sistema antigo de regime de responsabilidade (exige a compensação do valor do dano por parte do segurado). Concluiu que o sistema sem culpa atual em Quebec é ineficiente em comparação com o anterior, esse sistema apesar de ser atraente do ponto de vista de menores custos administrativos os resultados relativos a acidentes de trânsito não são favoráveis.

Hemenway (1992) realizou um estudo que buscava relacionar a tomada de risco físico com a tomada de risco medida pela compra de seguros, ele apresentou evidências sobre a possibilidade de haver entre essa relação o fator da seleção propícia que sugere que existem pessoas que evitam riscos, que tomam precauções e esses são mais propensos a adquirir um seguro de automóvel e aqueles que procuram algum tipo de risco tendem a não fazer nada, em sua pesquisa fatores como: uso de bebidas alcoólicas, cigarro, atividade física, fazer exames de rotina, como se alimentam e peso que possuem foram atributos analisados e os resultados passaram por um processo de regressão logística onde concluiu ter obtido resultados consistentes favoráveis à seleção propícia, mas que por causa de alguma limitações os resultados são sugestivos ao invés de conclusivos.

Blows et al. (2003) elaboraram um estudo na região de Auckland – Nova Zelândia que buscava relacionar o risco de se envolver em um acidente com a contratação do seguro do veículo, utilizando-se da metodologia de entrevista e preenchimento de formulário, onde os resultados foram analisados por regressão

logística e forneceram evidências de que no grupo estudado, os motoristas que não contratam seguro tem um risco aumentado de se ferir ou vir à óbito em um acidente do que àqueles que contratam. Os autores reconhecem que o mecanismo para uma relação causal entre o acidente de carro e a contratação de seguro não é clara, mas que os resultados possuem o mérito de importância, ele ainda reforça a explicação desses resultados concordando com a teoria da seleção propícia de Hemenway (1992).

O artigo de Adamos e Nathanail (2015) comenta sobre o crescimento da importância do seguro de veículo, informa ser uma consequência do crescimento do setor de veículos na Rússia que por sua vez se origina do crescimento da renda dos cidadãos e do seu bem-estar econômico. O estudo defende que as seguradoras devem ter uma abordagem orientada para o cliente, o que reduziria a ineficácia do segmento e cada cliente seria tratado de acordo com sua unicidade, isso traria maior transparência para o contratante, formaria clientes mais satisfeitos por estarem sendo cobrados de forma justa e essas medidas potencializariam a lucratividade da carteira de seguros.

Hsu et al. (2016) buscam em seu estudo realizado em Taiwan, relacionar os acidentes de trânsito, os seguros de veículos e a segurança geral do tráfego. Em sua primeira conclusão de acordo com o banco de dados estudado, verificou-se que os segurados que contratam níveis de seguro mais altos tendem a se envolver mais em acidentes de trânsito assim como os que possuem pobre histórico de condução, os resultados sugeriram a existência da seleção adversa (o segurado que contrata níveis maiores de seguros, toma tal decisão porque provavelmente irá se arriscar mais e provavelmente registrar algum tipo de sinistro). Em sua segunda parte do estudo onde inclui as informações de idade, gênero e ano do veículo para verificar o comportamento da relação de acidentes com a contratação de seguros, verifica-se que após inclusão desses dados as relações positivas originais tornam-se significativamente enfraquecidas.

As seguradoras costumam utilizar um sistema de bônus para oferecer desconto aos bons condutores que não tiverem registro de sinistro no último ano de seguro e reduzir o desconto para aqueles que registraram. No estudo de (GÓMEZ-DÉNIZ, 2016) ele diz que esse método é bom, mas pode parecer injusto pois não

diferencia os acidentes que geraram bastante prejuízos dos que foram mais razoáveis, penalizando os dois da mesma forma, sendo assim ele propõe um modelo estatístico que diferencia os dois diferentes tipos de sinistros: os que estão acima e abaixo de um valor considerado mediano. O estudo concluiu que essa modificação ofereceu credibilidade ao sistema satisfazendo as regras de transição desejáveis.

Ross et al. (2016) propuseram um estudo em que verifica-se que os jovens condutores que são tidos em estudos passados como a faixa etária de maior risco, tendem a se envolver mais em acidente quando estão acompanhados por passageiros pares, isto é, mostra-se que o jovem é altamente influenciável e a presença de um passageiro pode causar mais excitação ao dirigir, perdas de atenções momentâneas, ansiedade ao tomar decisões de emergências, fatores esses altamente indutáveis à um acidente, o estudo ainda identifica que os homens são mais adeptos a esses erros que as mulheres.

Stern et al. (2016) elaboraram um estudo que propunha um alerta para os condutores mais idosos, nos últimos 25 anos houve um aumento substancial do número de adultos licenciados nos EUA, estima-se que 20% da população terá mais de 65 anos até 2030, os motoristas mais velhos aumentam o risco de acidentes devido a doenças providas de suas idades elevadas, a condição neurológica é reduzida assim como alterações cognitivas ligeiras e a demência causada pelo Alzheimer, o estudo conclui que as doenças causam um maior risco ao condutor e que testes devem ser feitos com os segurados dessa faixa etária com o intuito de amenizar os riscos e se necessário haver uma intervenção do direito de dirigir caso os testes tiverem resultados abaixo do desejado.

2.2.1. Conclusão do Estado da Arte

De acordo com a revisão bibliográfica realizada foi possível verificar que as pesquisas relacionadas ao seguro de automóvel estão muito mais avançadas internacionalmente do que no Brasil, diversos estudos tentam de alguma forma relacionar a contratação de seguro com a quantidade de acidente que o condutor se envolve, os acidentes de trânsito aumentam a cada ano e se for provado por

estatísticas que as pessoas que contratam o seguro tendem a se envolver menos em acidente, torna-se um caso de saúde pública onde o Governo deve intervir seja com incentivos a contratação por parte dos segurados, ou com incentivos fiscais as seguradoras com o intuito de reduzirem os valores de apólices, ou com a realização de estudos que identifiquem a viabilidade de se tornar compulsório a contratação.

No âmbito judicial, verificam-se várias contestações sobre a criação de grupos de riscos de forma impessoal, demonstrando uma fragilidade do sistema e um possível pré-conceito que merece atenção do Governo, essas preconcepções geram insegurança por parte dos usuários e por muitas vezes levam a não contratação do seguro nos primeiros anos de direção devido ao alto valor cobrado. Ainda pode-se afirmar que grupos de baixa renda ficam limitados ao acesso ao benefício do seguro por geralmente se concentrarem em regiões de maior risco (periferias e centros) ou por não terem acesso a escolaridade.

Por fim, foi notada uma grande evolução nas pesquisas realizadas, a tentativa de tornar o cálculo mais pessoal já é muito se falado apesar de na prática ainda não ser comercializado, novas variáveis comportamentais têm sido estudadas para se medir o risco provável de cada pessoa se envolver em um evento de sinistro, substituindo as atuais que classificam os riscos por grupos pré-selecionados como faixas etárias, gênero, escolaridade, estado civil.

3. MATERIAIS E MÉTODOS

3.1. UTILIZAÇÃO DO KDD PARA OS DADOS SELECIONADOS

De acordo com o modelo proposto por Fayyad (1996), foram seguidos os passos do processo de KDD, descritos nas seções 3.1.1 à 3.1.5.

3.1.1. Seleção:

A Polícia rodoviária Federal – PRF fiscaliza cerca de 70 mil km de rodovias em todo o país, tendo o dever de combater a criminalidade, prestar apoio ao cidadão, fiscalizar, autuar e atender acidentes. Esses registros são realizados através do sistema BR-Brasil que está disponível desde 2007, justificando o motivo de esse ser o ano mais antigo a que se tem acesso (POLÍCIA FEDERAL, 2016)

Após análise dos dados divulgados pelo DPRF (2016) foi possível verificar planilhas contendo os dados de acidentes de trânsito registrados nos últimos 9 anos (o ano de 2016 foi utilizado porém vale ressaltar que só havia os dados dos 3 primeiros trimestres do ano), cada planilha é separada por ano e agrupada por pessoa, contendo todos registros de acidentes e para cada um continuam as variáveis que foram descritas na Tabela 4

Tabela 4 - Dicionário de variáveis

Id	Variável numérica que representa o acidente
Data_inversa	Data que ocorreu o acidente
Dia_semana	Dia da semana: segunda-feira, terça-feira...
Horário	Horário de ocorrência
UF	Unidade Federativa: SP, ES, RJ, MG
BR	Nome da rodovia
KM	Quilometragem que ocorreu o acidente (trecho)
Município	Nome do município
Causa_acidente	Causa do acidente: falta de atenção, alta velocidade, ingestão de álcool, animais na pista, defeito mecânico em veículo, defeito na via, desobediência a sinalização, dormindo, não guardar distância segura, ultrapassagem indevida, outras e NULL.
Tipo_acidente	Como ocorreu: atropelamento de animal, atropelamento de pessoa, capotamento, colisão com objeto fixo, colisão com objeto móvel, colisão frontal, colisão transversal, colisão traseira, danos eventuais, derramamento de carga, incêndio, saída de pista, tombamento, queda de motocicleta/bicicleta/veículo e NULL.
Classificação_acidente	Gravidade do acidente: com vítima fatais, vítimas feridas, sem vítima, ignorado e NULL.
Fase_dia	Momento do dia: amanhecer, anoitecer, plena noite, pleno dia e NULL
Sentido_via	Sentido da via: crescente ou decrescente
Condição_meteorológica	No momento da colisão estava: céu claro, chuva, granizo, neve, nevoeiro/neblina, nublado, sol, vento, ignorado, NULL, vazias.
Tipo_pista	Tipo da pista: simples, dupla ou mista.
Traçado_via	Descrição: reta, curva, cruzamento.
Uso_solo	Tipo do solo: urbano ou rural.
Id_veiculo	Variável numérica que identifica os carros.
Tipo_veiculo	Tipo do veículo: automóvel, motocicleta, caminhão, automóvel, bicicleta, bonde/trem, caminhão, caminhão-tanque, caminhão-tractor, caminhonete, camioneta, carroça, carro-de-mão, charrete, ciclomotor, motocicleta, micro-ônibus, motoneta, motor-casa, ônibus, quadriciclo, reboque, semi-reboque, trator de esteiras, trator de rodas, trator misto, triciclo, utilitário, ignorado, NULL, vazias.
Marca	Marca e modelo do veículo, não convém escrever todos os registros por se tratar de mais de 300 diferentes modelos descritos
Ano_fabricação_veiculos	Ano em que o veículo foi fabricado
Pesid	Variável numérica que identifica as pessoas
Tipo_envolvido	Participação no evento: condutor, passageiro, pedestre.
Estado_físico	Condição do envolvido após acidente: ferido grave, ferido leve, ileso, morto, ignorado.
Idade	Idade do envolvido, onde (-1) significa que não foi possível coletar tal dado.
Gênero	Gênero do envolvido: masculino, feminino, não identificado
Nacionalidade	Nacionalidade do envolvido
Naturalidade	Naturalidade do envolvido

Fonte: Elaborado pelo Autor (2017)

3.1.2. Pré-Processamento

Os dados precisaram ser limpos antes de iniciar o processo de mineração de dados através do WEKA, para isso, algumas variáveis que não agregariam valor a pesquisa tiveram que ser retiradas, sendo elas: id, data_inversa, dia_semana, horário, BR, KM, município, sentido_via, id_veiculo, pesid, nacionalidade e naturalidade.

Optou-se por excluir todos os dados que tivessem a seguinte denominação: “outras”, “null”, “ignorado”, “não identificado”, “vazias”, “xxx”, “-1”, “NA”, “invalido”. No variável tipo_envolvido, foram mantidos apenas os condutores pois são o foco deste trabalho.

O algoritmo K-means não faz leitura de dados numéricos, para isso foi necessário acrescentar aos atributos “Idade” e “Ano fabricação” as letras i e a respectivamente que foram selecionadas de forma aleatória, repetindo o processo em todos os registros de acidentes.

3.1.3. Transformação

A próxima etapa foi criar uma nova planilha do Excel, salvar em formato CSV e agrupar os dados por estado ao invés de ano, separando o gênero masculino do feminino, isso se torna necessário pois o arquivo de CSV não faz leitura de filtros, logo, para cada grupo de dados que precisava-se efetuar leitura no WEKA era necessária uma nova planilha, ao final obteve-se 8 novas planilhas.

3.1.4. Mineração de dados

O algoritmo K-means foi proposto por JMacQueen em 1967, é um clássico algoritmo de agrupamento de dados utilizado de forma ampla na pesquisa científica (YANG; DENG, 2010).

O método de mineração de dados, é primordial para extrair conhecimento de uma grande base de dados, várias técnicas são aplicadas para se obter esse

resultado esperado, entre elas encontra-se o agrupamento de dados, o K-means é um versátil, veloz e robusto método (SABITHA; KARTHIK, 2015).

O processo de clusterização do Simple K-means foi aplicado ao banco de dados pelo software WEKA 3.6 para ser verificado qual o padrão de dados mais relevante para cada estado e gênero.

3.2. GRÁFICOS DEMONSTRATIVOS DA SUSEP

O Sistema Autoseg apresenta as informações de acordo com o grupamento escolhido pelo usuário, que representa o conjunto de apólices incluídas nesse determinado grupo de pesquisas, optou-se então por pesquisar: Categoria Tarifária/ Região/ Modelo/ Ano/ Gênero/ Idade, onde foram aplicados os seguintes filtros (SUSEP, 2016):

(I). Categoria Tarifária:

- Veículo de passeio nacional;
- Veículo de passeio importado;
- Veículo modelo Pick-up (nacional e importado);
- Veículo Utilitário (nacional e importado).

(II). Região: as regiões estão separadas não só por Estado mas também algumas especificações de localidades, quando no trabalho for citado RJ estarão englobadas todas as regiões (interior e Metropolitanas), SP (Grande Campinas, Litoral Norte e Baixada Santista, metropolitanas, Ribeirão Preto e demais municípios de Campinas, Vale do Paraíba e Ribeira), MG (metropolitanas BH – Centro, Oeste-Zona Mata-C. Vertentes, Sul, Triângulo Mineiro, Vale do Aço-Norte-Vale Jequitinhonha), ES (interior e metropolitana).

(III). Grupos de modelos: Todos.

(A). Ano modelo: Todos.

(B). *Gênero do condutor*: Masculino e Feminino.

(C). *Faixa etária*: Entre 18 e 25 anos, entre 26 e 35 anos, entre 36 e 45 anos, entre 46 e 55 anos, maior que 55 anos.

(IV). **Período de pesquisa**: 01/07/2006 à 31/12/2015

Esses dados foram exportados para uma planilha de Excel onde posteriormente foram transformados em gráficos para melhor visualização. As variáveis compostas nos dados gerados foram aqui nomeadas e explicadas por objetivo conceitual.

(1). Prêmio médio (R\$): o prêmio médio representa a média dos prêmios das apólices incluídas no grupamento, ponderada pela exposição de cada uma delas.

(2). IS Média (R\$): A Importância Segurada Média (IS Média) representa a média das IS's das apólices incluídas no grupamento, ponderada pela exposição de cada uma delas (ver conceito de exposição abaixo), essa variável não foi utilizada nas tabelas de resultado

(3). Expostos: O conceito de exposição leva em conta o tempo em que cada apólice esteve vigente, dentro da janela de observação, que é o período semestral abrangido em cada atualização do Autoseg. Desta forma, o número de expostos, apurado para um período anual, representa o melhor estimador disponível para a quantidade de veículos segurados.

(4). Frequência: A Frequência é a quantidade de sinistros de incêndio, roubo, colisão e outras causas.

(5). Frequência (outras): Frequência outras engloba os sinistros não incluídos nos dados de incêndio, roubo, ou colisão, ou seja, basicamente assistência 24 horas e outras coberturas como vidros, blindagem, equipamentos acessórios entre outros.

(6). GM (R\$): variável criada que soma todas indenizações (Indenização incêndio roubo (R\$), Indenização Colisão (R\$), Indenização outras (R\$) e divide pelo somatório das Frequências (Frequência Incêndio Roubo, Frequência Colisão, Frequência Outras).

(7). TA (%): Variável criada que demonstra de forma percentual quantos indivíduos registraram um sinistro em razão dos que contrataram, esse valor é obtido a partir do somatório de todas as Frequências (Frequência Incêndio Roubo, Frequência Colisão, Frequência Outras) dividido pelo somatório do número de expostos do determinado filtro.

(8). Prejuízo (R\$): Para poder ter uma melhor noção da média de gastos que uma seguradora possui com cada faixa etária, foi criada essa variável que é o resultado da multiplicação de (TA) por (GM). Essa variável possui o intuito de evitar análises errôneas dos dados da SUSEP, onde pode observar-se que em algumas faixas etárias o TA é menor, porém o GM é maior e a análise única de um desses dados pode gerar falsas conclusões.

(9). Amplitude (%): É a relação percentual de redução dos valores de Prejuízo de todas as faixas etárias em relação a de 18 a 25 anos, busca-se verificar com ela o quanto mais barato é o custo de um acidente para as demais faixas etárias. Nos casos onde o resultado for negativo, significa que para tal faixa etária é mais caro o custo com um sinistro do que para a supracitada.

3.3. CÁLCULO DO PRÊMIO PAGO PARA CADA FAIXA ETÁRIA

Oito personagens foram criados com as mesmas características básicas para que o foco do problema não sofra com as variáveis desprezíveis para essa pesquisa, como são quatro Estados dentro da Região Sudeste, em cada Estado foram usados dois perfis (masculino e feminino) se diferenciando apenas pela idade, buscando obter a diferença de valor com foco na faixa etária. O cálculo foi efetuado de forma online no site: www.autocompara.com.br que fornece o cálculo de 4 seguradoras que mantém esses dados públicos: Sul América, Tokio Marine Seguradora, HDI Seguros e Allianz.

O cálculo da data de nascimento do condutor foi efetuado da seguinte maneira: o dia e mês colocou-se o mesmo da data em que foi efetuado o cálculo (12 de Outubro) e o ano foi uma média feita de cada faixa etária:

Faixa etária 1 - $(18+25) / 2 = 22$ anos, nascido em 1994;

Faixa etária 2 - $(26+35) / 2 = 31$ anos, nascido em 1985;

Faixa etária 3 - $(36+45) / 2 = 41$ anos, nascido em 1975;

Faixa etária 4 - $(36+45) / 2 = 51$ anos, nascido em 1965;

Faixa etária 5 - Maior que 55 = 60 anos, nascido em 1956;

3.4. CARACTERÍSTICAS QUE FORAM MANTIDAS EM TODOS OS PERFIS:

(A). Veículo: Volkswagen Gol Confortiline 1.0 G6 8v flex

(B). Ano: 2016/2016

(C). Veículo zero KM: Sim;

(D). Extensão de garantia: Não

(F). Chassi remarcado: Não;

(D). Kit gás: Não;

(F). Veículo blindado: Não;

(G). Veículo possui rastreador ou bloqueador: Não;

(H). Caso oferecido, deseja instalar um sistema de segurança: Não;

(I). Deseja incluir acessórios adicionais (som/vídeo): Não;

(J). Endereço: mudou conforme a mudança de Estado, porém todos foram feitos na região central das capitais nos CEPs: 28030-335 RJ, 39900-000 MG, 18133-445 SP, 29075-910 ES;

(K). Veículo utilizado para: Exclusivamente lazer;

(L). Tipo de residência: Casa;

(M). Possui garagem fechada: Na residência;

(N). Veículo utilizado por deficiente físico: Não;

(O). Tipo de utilização do veículo: Particular;

(P). Já possui seguro: Não;

(Q). Houve sinistro no último ano: Não;

(R). CPF: Foi escolhido um CPF válido, de forma aleatória.

(S). Gênero: Masculino/Feminino;

(T). Data de nascimento: foram utilizadas 5 datas (12/10/1994, 12/10/1985, 12/10/1975, 12/10/1965, 12/10/1956);

(U). Estado civil: Solteiro;

(V). Cobertura para outros condutores entre 17 e 25 anos: Não;

(X). Segurado é o proprietário do veículo: Sim;

(Z). Principal condutor é o mesmo que o proprietário: Sim;

4. RESULTADOS

4.1. RESULTADO DO SIMPLEKMEANS NO SOFTWARE WEKA

As tabelas aqui apresentadas foram adaptadas do resultado obtido pelo Software WEKA 3.6, para possibilitar a comparação entre eles, para cada grupo de atributos (gênero e estado) foi gerado um novo padrão através do SimpleKmeans. Para análise de resultados foram levados em consideração os dados “Full Data”.

Na Tabela 5 é possível identificar os principais fatores de risco no estado de São Paulo e gênero masculino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se teve registro de vítimas, ocorreram em pleno dia com céu claro, pista dupla em retas, trecho urbano, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e o condutor com idade de 30 anos saiu ileso.

Tabela 5 - Resultados obtidos do Estado de São Paulo (SP): gênero masculino

ATRIBUTOS	FULL DATA (93578.0)
Estado	SP
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Dupla
Traçado via	Reta
Uso do solo	Urbano
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Masculino
Idade	30 i
Ano fabricação	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 6 é possível identificar os principais fatores de risco no estado de São Paulo e gênero feminino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista dupla em retas, trecho urbano, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e a condutora com idade de 26 anos saiu ileso.

Tabela 6 - Resultado obtido do Estado de São Paulo (SP): gênero feminino

ATRIBUTOS	FULL DATA (10359.0)
Estado	SP
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Dupla
Traçado via	Reta
Uso do solo	Urbano
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Feminino
Idade	26 i
Ano Fabricação	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 7 é possível identificar os principais fatores de risco no estado de Rio de Janeiro e gênero masculino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista simples em retas, trecho urbano, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e o condutor com idade de 31 anos saiu ileso.

Tabela 7 - Resultado obtido do Estado do Rio de Janeiro (RJ): gênero masculino

ATRIBUTOS	FULL DATA (81674.0)
Estado	RJ
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Simples
Traçado via	Reta
Uso do solo	Urbano
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Masculino
Idade	31 i
Ano	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 8 é possível identificar os principais fatores de risco no estado de Rio de Janeiro e gênero feminino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista simples em retas, trecho urbano, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e o condutor com idade de 28 anos saiu ileso.

Tabela 8 - Resultado obtido do Estado do Rio de Janeiro (RJ): gênero feminino

ATRIBUTOS	FULL DATA (7990.0)
Estado	RJ
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Simplex
Traçado via	Reta
Uso do solo	Urbano
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Feminino
Idade	28 i
Ano	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 9 é possível identificar os principais fatores de risco no estado de Minas Gerais e gênero masculino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista simples em retas, trecho rural, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e o condutor com idade de 31 anos saiu ileso.

Tabela 9 - Resultado obtido do Estado de Minas Gerais (MG): gênero masculino

ATRIBUTOS	FULL DATA (188282.0)
Estado	MG
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem Vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu Claro
Tipo de pista	Simplex
Traçado via	Reta
Uso do solo	Rural
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Masculino
Idade	31 i
Ano	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 10 é possível identificar os principais fatores de risco no estado de Minas Gerais e gênero feminino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista simples em retas, trecho rural, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e a condutora com idade de 30 anos saiu ileso.

Tabela 10 - Resultado obtido do Estado de Minas Gerais (MG): gênero Feminino

ATRIBUTOS	FULL DATA (14323.0)
Estado	MG
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Simples
Traçado via	Reta
Uso do solo	Rural
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Feminino
Idade	30 i
Ano	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 11 é possível identificar os principais fatores de risco no estado do Espírito Santo e gênero masculino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista simples em retas, trecho urbano, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e o condutor com idade de 30 anos saiu ileso.

Tabela 11 - Resultado obtido do Estado do Espírito Santo (ES): gênero Masculino

ATRIBUTOS	FULL DATA (79869.0)
Estado	ES
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Simples
Traçado via	Reta
Uso do solo	Urbano
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Masculino
Idade	30 i
Ano	2008 a

Fonte: Elaborado pelo Autor (2017).

Na Tabela 12 é possível identificar os principais fatores de risco no estado do Espírito Santo e gênero feminino, sendo eles: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se obteve registro de vítimas, ocorreram em pleno dia com céu claro, pista simples em retas, trecho urbano, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008 e o condutor com idade de 34 anos saiu ileso.

Tabela 12 - Resultado obtido do Estado do Espírito Santo (ES): gênero Feminino

ATRIBUTOS	FULL DATA (7280.0)
Estado	ES
Causa do acidente	Falta de atenção
Tipo acidente	Colisão traseira
Classificação do acidente	Sem vítimas
Fase do dia	Pleno dia
Condição meteorológica	Céu claro
Tipo de pista	Simple
Traçado via	Reta
Uso do solo	Urbano
Tipo do veículo	Automóvel
Marca veículo	VW/Gol 1.0
Tipo envolvido	Condutor
Estado físico	Ileso
Gênero	Feminino
Idade	34 i
Ano fabricação	2008 a

Fonte: Elaborado pelo Autor (2017).

4.2. RESULTADOS OBTIDOS A PARTIR DA BASE DE DADOS DA SUSEP

Os Gráficos 4 ao gráfico 15 desse capítulo possuem o objetivo de demonstrar o comportamento que cada faixa etária do gênero masculino e feminino tiveram em cada Estado da região Sudeste e assim foi possível classificar com base nos registro da SUSEP qual é a ordem de risco do maior para o menor das faixas etárias dos condutores levando em consideração os respectivos prejuízos causados por elas, as siglas foram definidas no item 3.2 na p. 44.

No Gráfico 4 foi possível identificar uma linearidade de dados, onde a medida que a faixa etária sobe, o GM é reduzido junto com o PREJUÍZO e o TA, o que mostra um menor risco para o perfil de pessoas com a idade mais elevada. A amplitude que é inversamente proporcional ao PREJUÍZO aumenta atingindo um

valor de 54,98% o que significa que esse percentual é quanto menos PREJUÍZO essa faixa etária ofereceu em relação a de 18 a 25 anos. Ordem de risco do maior para a menor das faixas: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 36 a 45 anos” – “entre 46 a 55 anos” – “> que 55 anos”.

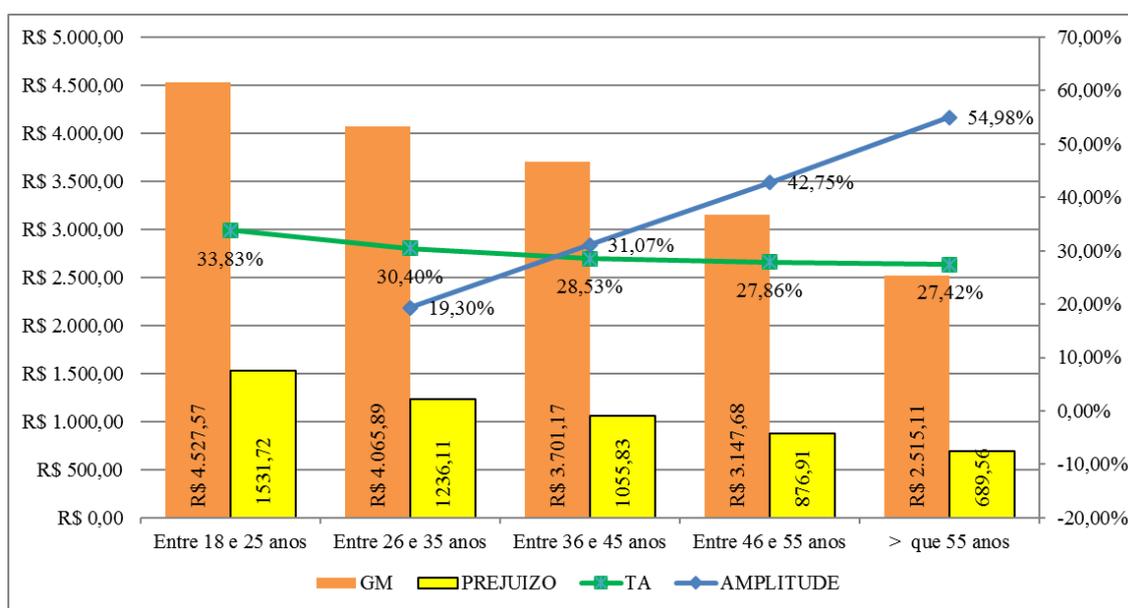


Gráfico 4. Resultado da análise para o Estado do Rio de Janeiro (RJ) de pessoas do gênero masculino.

Fonte: Elaborado pelo Autor (2017).

No Gráfico 5 diferentemente do masculino, a medida que a faixa etária se eleva registram mais aberturas de sinistros conforme se observa o crescimento do TA, porém esses sinistros são de menor escala e possuem menor GM, portanto conforme observado o PREJUÍZO é decrescente e a ordem de risco para as seguradoras do maior para o menor fica: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 36 a 45 anos” – “entre 46 a 55 anos” – “> que 55 anos”.

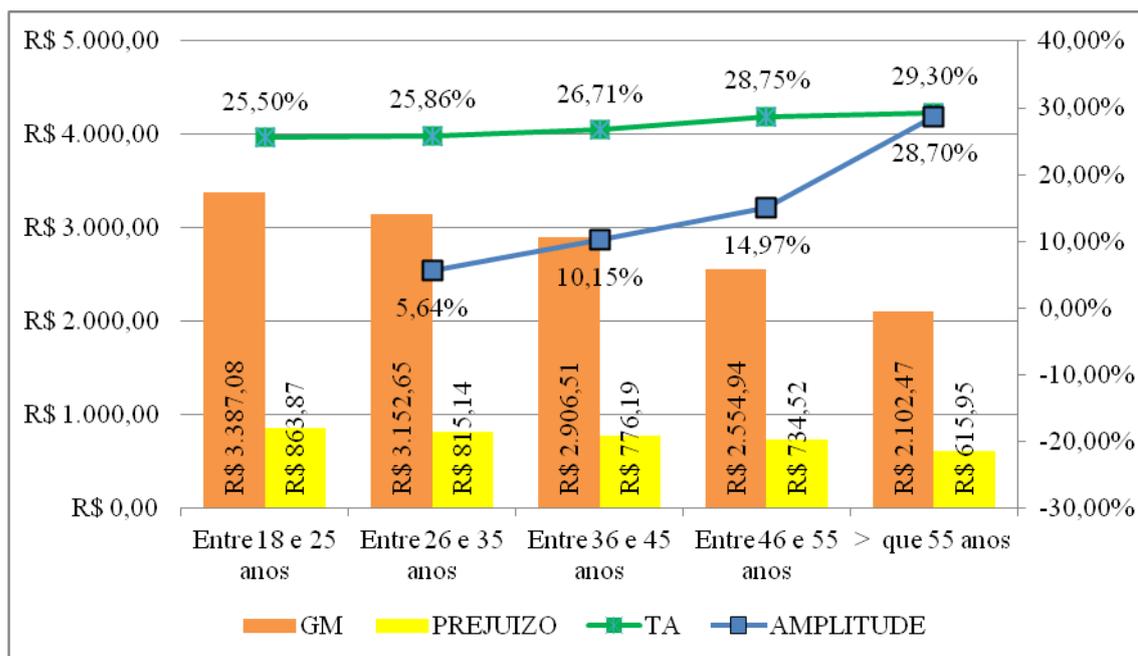


Gráfico 5. Resultado da análise para o Estado do Rio de Janeiro (RJ) de pessoas do gênero feminino.

Fonte: Elaborado pelo Autor (2017).

No Gráfico 6 observa-se que o GM foi reduzindo a medida que a faixa etária aumentou, o TA teve seu maior registro na faixa de 18 a 25 anos e reduziu até a de 36 a 45 anos onde voltou a aumentar nas duas próximas, mas esses dados não foram suficientemente determinantes ao ponto de modificar a ordem de risco do maior para a menor das faixas que manteve em: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 36 a 45 anos” – “entre 46 a 55 anos” – “> que 55 anos” justificado pelo PREJUÍZO que seguiu essa ordem.

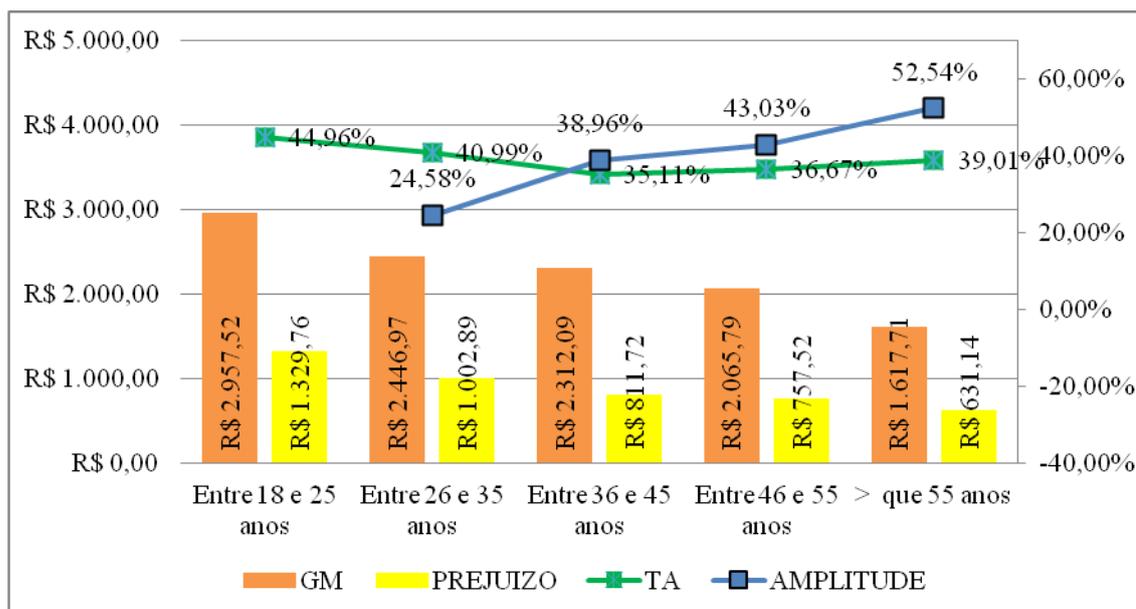


Gráfico 6. Resultado da análise para o Estado de São Paulo (SP) de pessoas do gênero Masculino.
Fonte: Elaborado pelo Autor (2017).

No gráfico 7 o comportamento do TA chama bastante atenção nesse gráfico, pois demonstra que a medida que a idade das mulheres aumenta, elas se envolveram mais em acidentes atingindo um percentual de 52,32% que significa que em média a cada duas pessoas que contrataram seguro dessa faixa etária, uma registrou sinistro. Em contrapartida os valores de GM das mulheres com mais de 55 anos é menor que a metade das que estão entre 18 a 25 anos. Ao multiplicar os valores de GM por TA obtemos o PREJUÍZO e então a ordem de risco desses valores do maior para o menor fica: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 46 a 55 anos” - “entre 36 a 45 anos” - “> que 55 anos”.

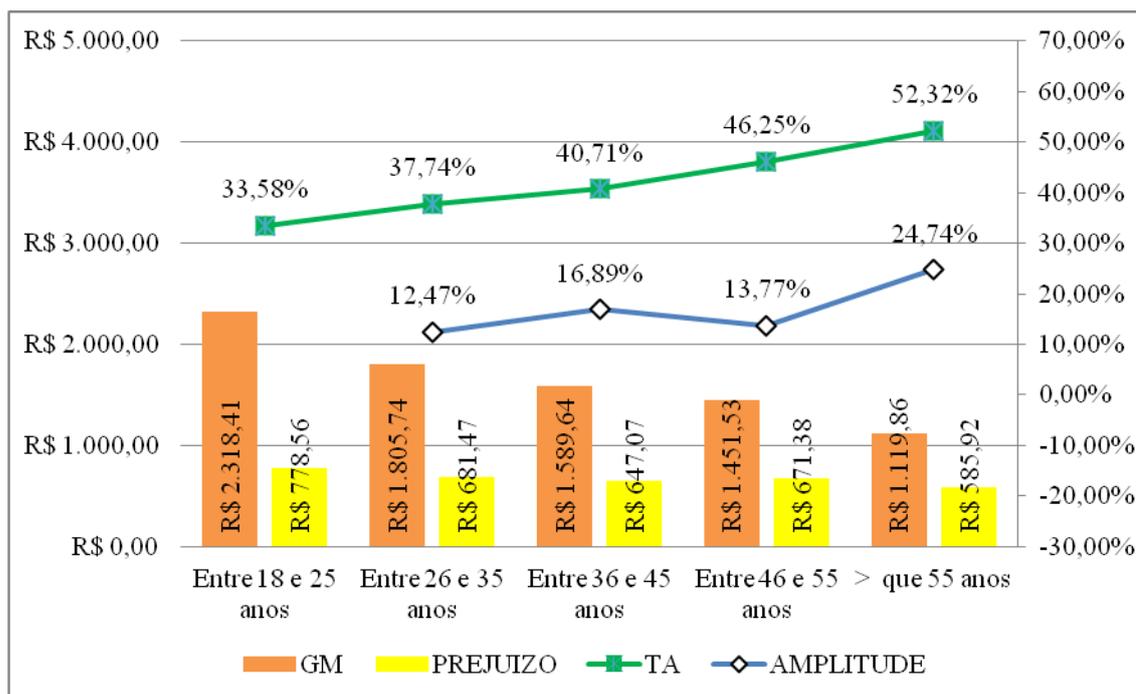


Gráfico 7. Resultado da análise para o Estado de São Paulo (SP) de pessoas do gênero Feminino.
Fonte: Elaborado pelo Autor (2017).

No Gráfico 8 observa-se que o GM foi oscilatório nas faixas, não seguindo uma ordem crescente ou decrescente, seu maior valor foi registrado na faixa de 18 a 25 anos e o menor na de maior que 55 anos, assim como o TA que não teve uma ordem definida, onde registra seu maior valor na faixa etária de 26 a 35 anos e menor na de 46 a 55 anos. Ao multiplicar essas duas variáveis, obtêm-se o valor do PREJUÍZO que leva a concluir que a ordem de risco do maior para o menor fica: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 36 a 45 anos” – “entre 46 a 55 anos” – “> que 55 anos”.

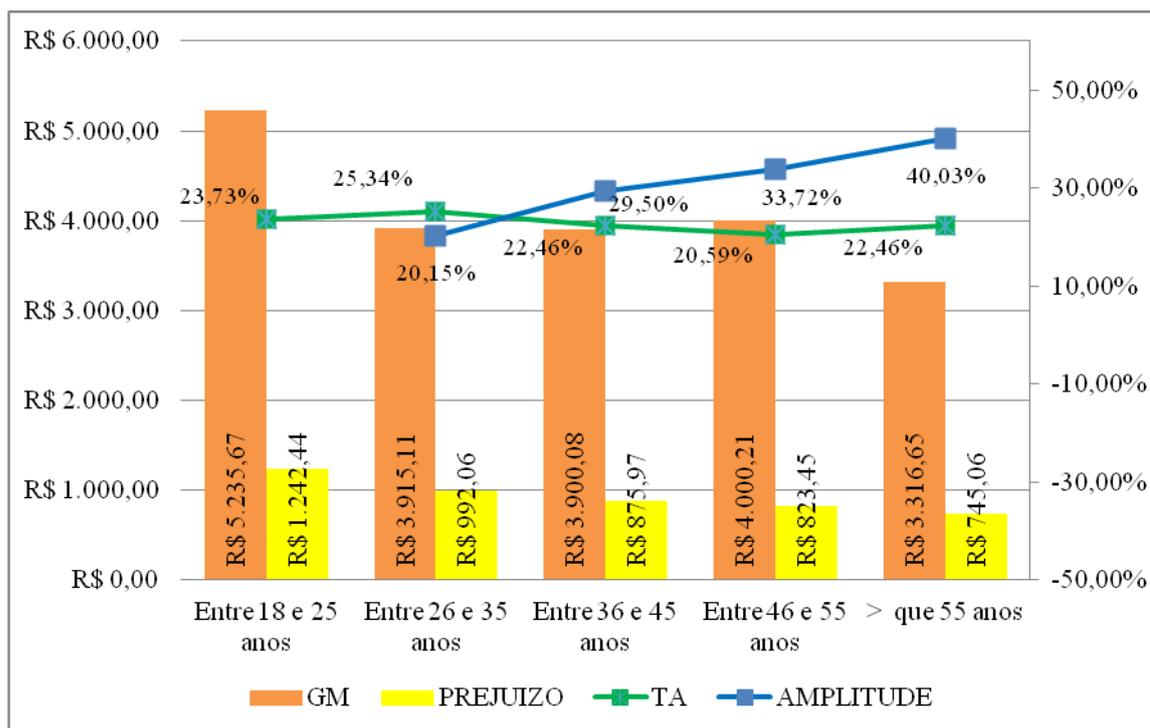


Gráfico 8. Resultado da análise para o Estado do Espírito Santo (ES) de pessoas do gênero Masculino.

Fonte: Elaborado pelo Autor (2017).

No Gráfico 9 o que chama mais atenção é a amplitude que aparece com valores negativos, a amplitude como é o coeficiente de redução dos valores de PREJUÍZO de todas as faixas etárias em relação a faixa de 18 a 25 anos, aparece negativo quando alguma dessas faixas apresenta um valor de PREJUÍZO superior, sendo assim, verifica-se que o maior valor de prejuízo encontra-se na faixa de 46 a 55 anos, outro dado que se destaca é o TA que mostra que a medida que eleva-se a faixa etária a taxa de abertura de sinistro também aumenta, gerando-se então a seguinte ordem de risco do maior para o menor: “entre 46 a 55 anos” – “entre 26 a 35 anos” - “entre 18 a 25 anos”– “entre 36 a 45 anos” — “> que 55 anos”.

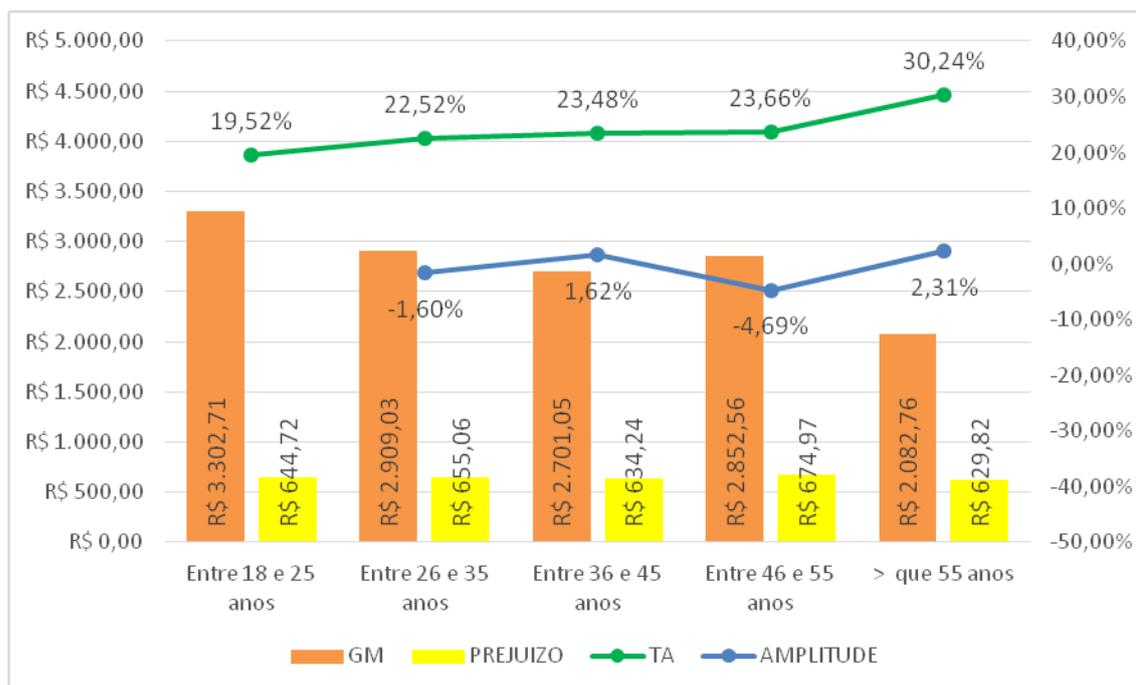


Gráfico 9. Resultado da análise para o Estado do Espírito Santo (ES) de pessoas do gênero Feminino.

Fonte: Elaborado pelo Autor (2017).

No Gráfico 10 observa-se o GM e o TA mais elevado na faixa etária de 18 a 25 anos, isso mostra que essa faixa é de grande risco nesse estado para condutores do gênero masculino, quando analisado a Amplitude é possível observar que a faixa etária de maior que 55 anos teve uma média de PREJUÍZO 52,17% menor que a faixa 18 a 25 anos, uma grande diferença percentual. Destaca-se também os altos valores apresentados pelo GM em todas as faixas e o TA decrescente a medida que se avança a idade do condutor, a ordem de risco do maior para o menor fica: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 46 a 55 anos” - “entre 36 a 45 anos” - “> que 55 anos”.

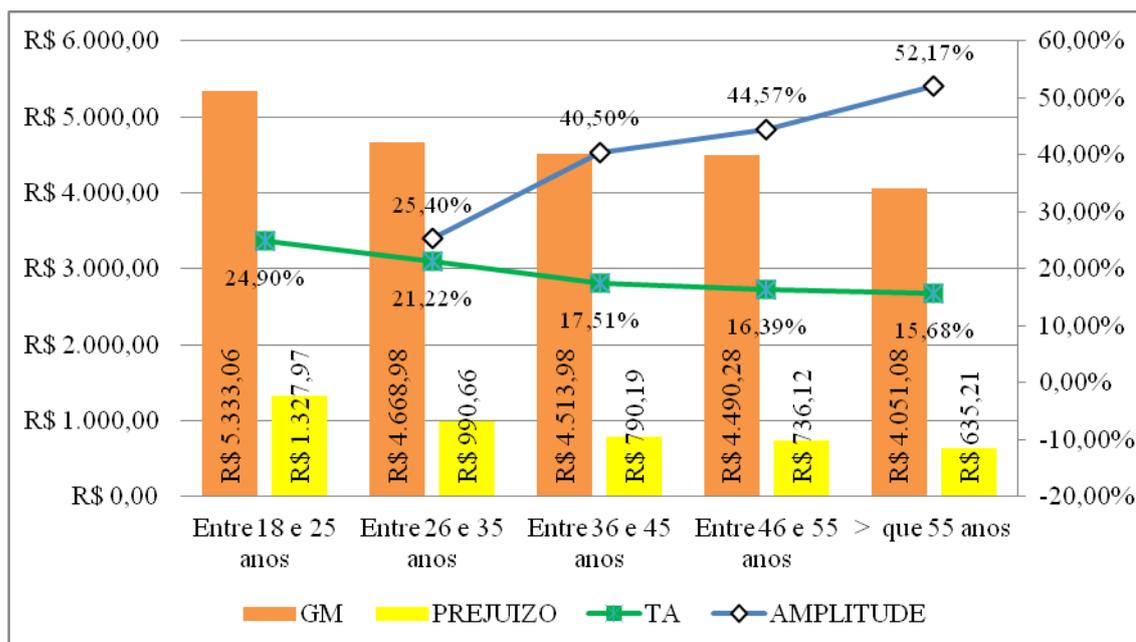


Gráfico 10. Resultado da análise para o Estado de Minas Gerais (MG) de pessoas do gênero Masculino.

Fonte: Elaborado pelo Autor (2017).

No Gráfico 11 é possível se verificar uma oscilação no GM, o maior valor é registrado na faixa etária de 18 a 25 anos assim como o maior TA, o menor PREJUÍZO está na faixa etária de maiores que 55 anos assim como o menor GM, o TA mais baixo aparece na faixa etária de 36 a 45 anos e então a ordem de risco do maior para o menor das faixas fica: “entre 18 a 25 anos” - “entre 46 a 55 anos” - “entre 26 a 35 anos” -- “entre 36 a 45 anos” - “> que 55 anos”.

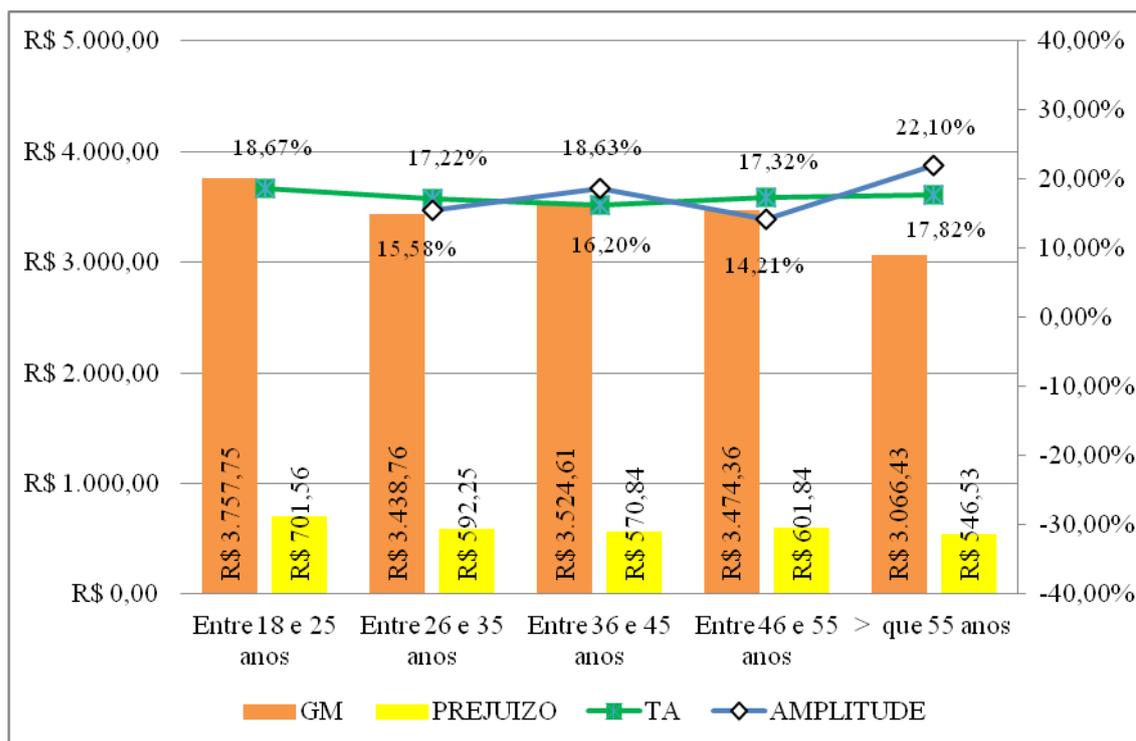


Gráfico 11. Resultado da análise para o Estado de Minas Gerais (MG) de pessoas do gênero Feminino.

Fonte: Elaborado pelo Autor (2017).

No Gráfico 12 possibilita verificar que os maiores gastos com os sinistros estão no gênero masculino na faixa etária de 18 a 25 anos nos Estados de Minas Gerais e Espírito Santo respectivamente e o menor valor médio foi registrado em São Paulo para pessoas com mais de 55 anos do gênero feminino. Os Gastos Médios dos homens em uma ocorrência de sinistro foram superiores em todas faixas etárias e Estados estudados com relação às mulheres. Verifica-se também que apenas os Estados de RJ e SP obtiveram valores decrescentes consecutivamente no gênero masculino e feminino quando foi elevada a idade de análise.

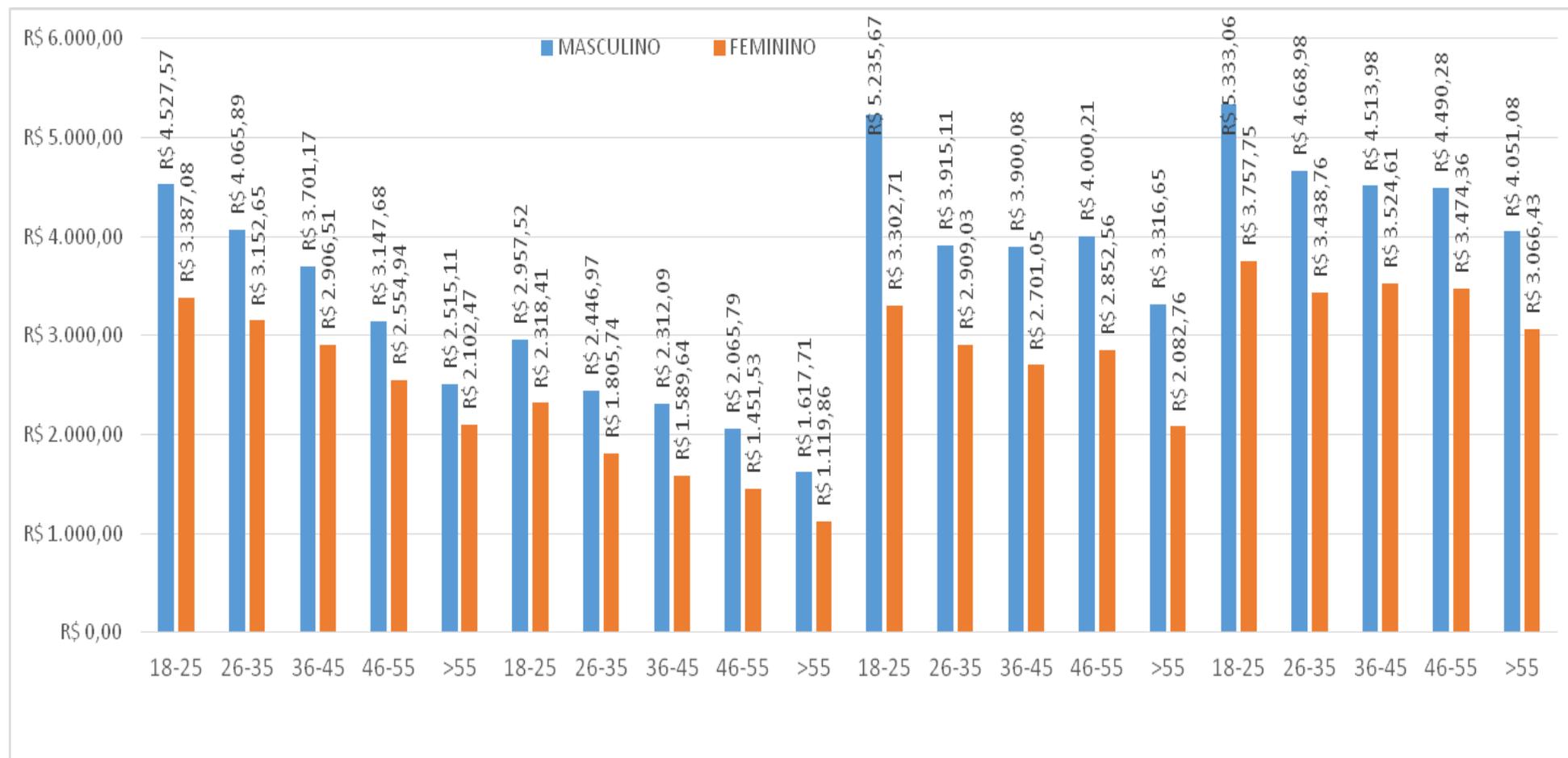


Gráfico 12. Análise por Gasto Médio (GM).

Fonte: Elaborado pelo Autor (2017).

No gráfico 13 o maior valor de TA foi registrado no Estado de São Paulo para as mulheres com mais de 55 anos que tiveram maior percentual de abertura de sinistros e os que obtiveram o menor valor de TA foram os homens de MG. No Estado do RJ quanto maior a idade dos homens maior foi o registro de sinistros e com as mulheres ocorreu o oposto, quanto maior a idade, menos registros tiveram; No Estado de SP os homens apresentaram dados voláteis e as mulheres registram mais sinistros a medida que a idade foi elevada; No ES ocorre situação similar a de SP, onde os homens apresentarão volatilidade nos resultados e as mulheres TA crescente a medida que se eleva a idade. No estado de MG tanto os homens quanto as mulheres apresentaram volatilidade

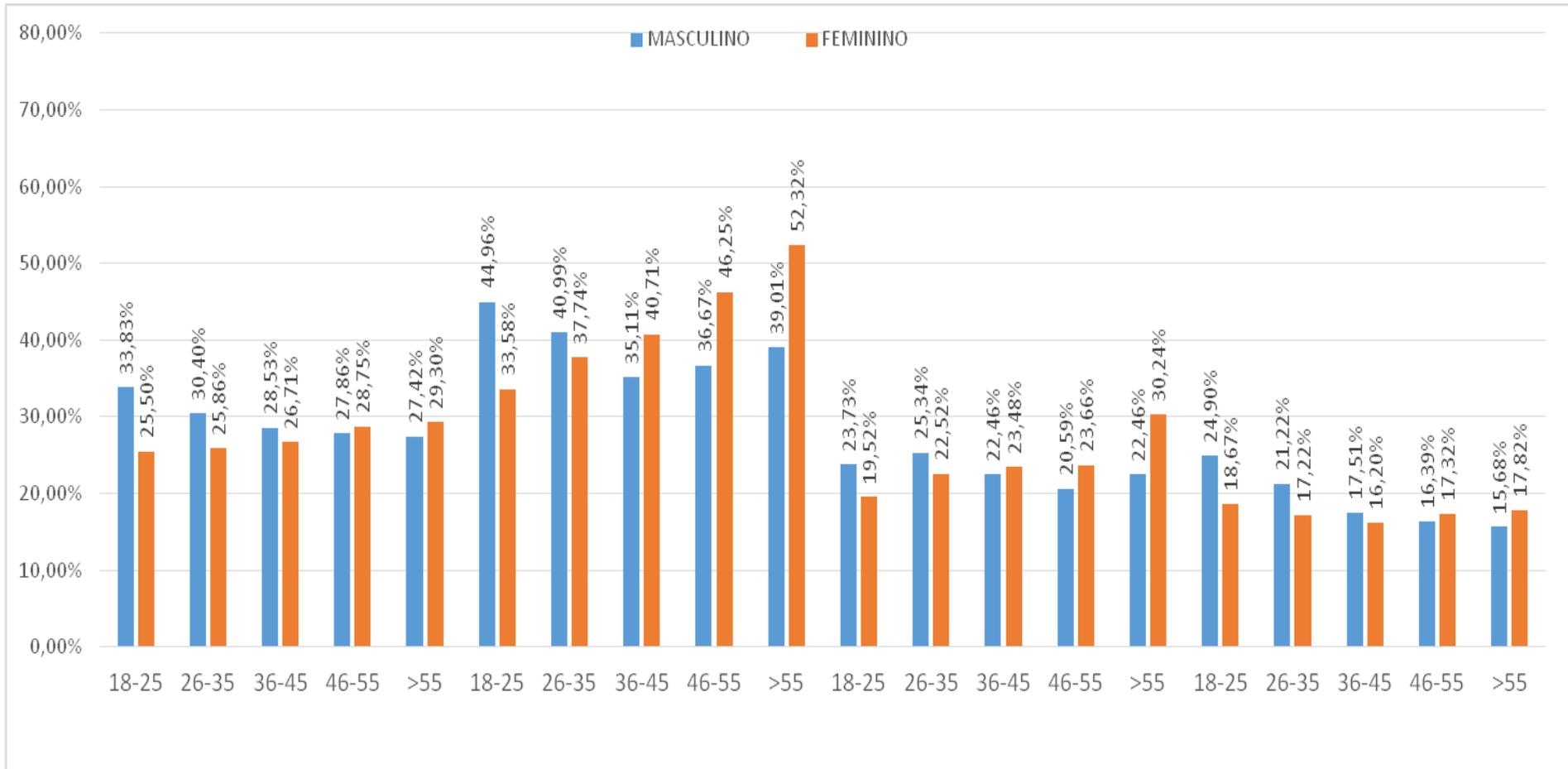


Gráfico 13. Análise por Taxa de Acidente (TA)
 Fonte: Elaborado pelo Autor (2017).

No gráfico 14 verifica-se que o maior PREJUÍZO foi registrado pela faixa etária de 18 a 25 anos do gênero masculino no estado do Rio de Janeiro e o menor foram as mulheres com mais de 55 anos do estado de Minas Gerais, em todos os casos analisados os homens oferecem maior PREJUÍZO às seguradoras do que as mulheres. O gênero masculino em todos estados teve a variável decrescente à medida que se avança a idade, ou seja, quanto mais jovem maior o PREJUÍZO, já as mulheres só obtiveram resultados decrescentes no estado do Rio de Janeiro, nos demais oscilatório.

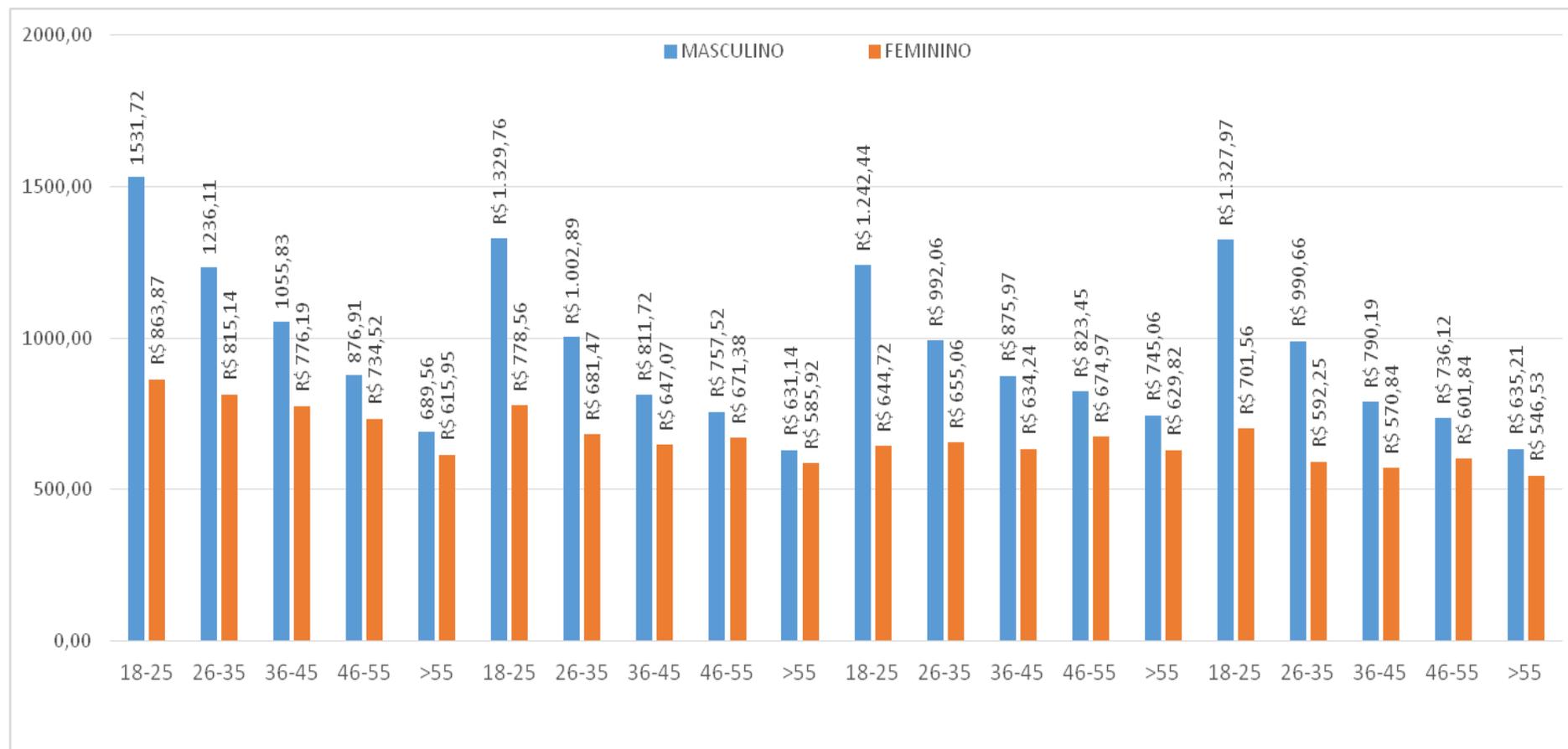


Gráfico 14. Análise por Prejuízo
 Fonte: Elaborado pelo Autor (2017).

No Gráfico 15 observa-se valores altos de amplitude dos valores quando observado o gênero masculino, valores estes bem superiores em todos estados em comparativo com as mulheres, no Rio de Janeiro registrou-se 54,98% menos de PREJUÍZO para a faixa etária dos homens com mais de 55 anos do que os de faixa de 18 a 25 anos, assim também é possível identificar que as mulheres possuem essa variância menor e ainda obteve-se valor negativo no estado de Minas Gerais que significa que (se essa variável é o percentual de redução de PREJUÍZO de todas faixas etárias em relação a de 18 a 25 anos), essas faixas então apresentam valores menores de PREJUÍZO que a de 18 a 25 anos.

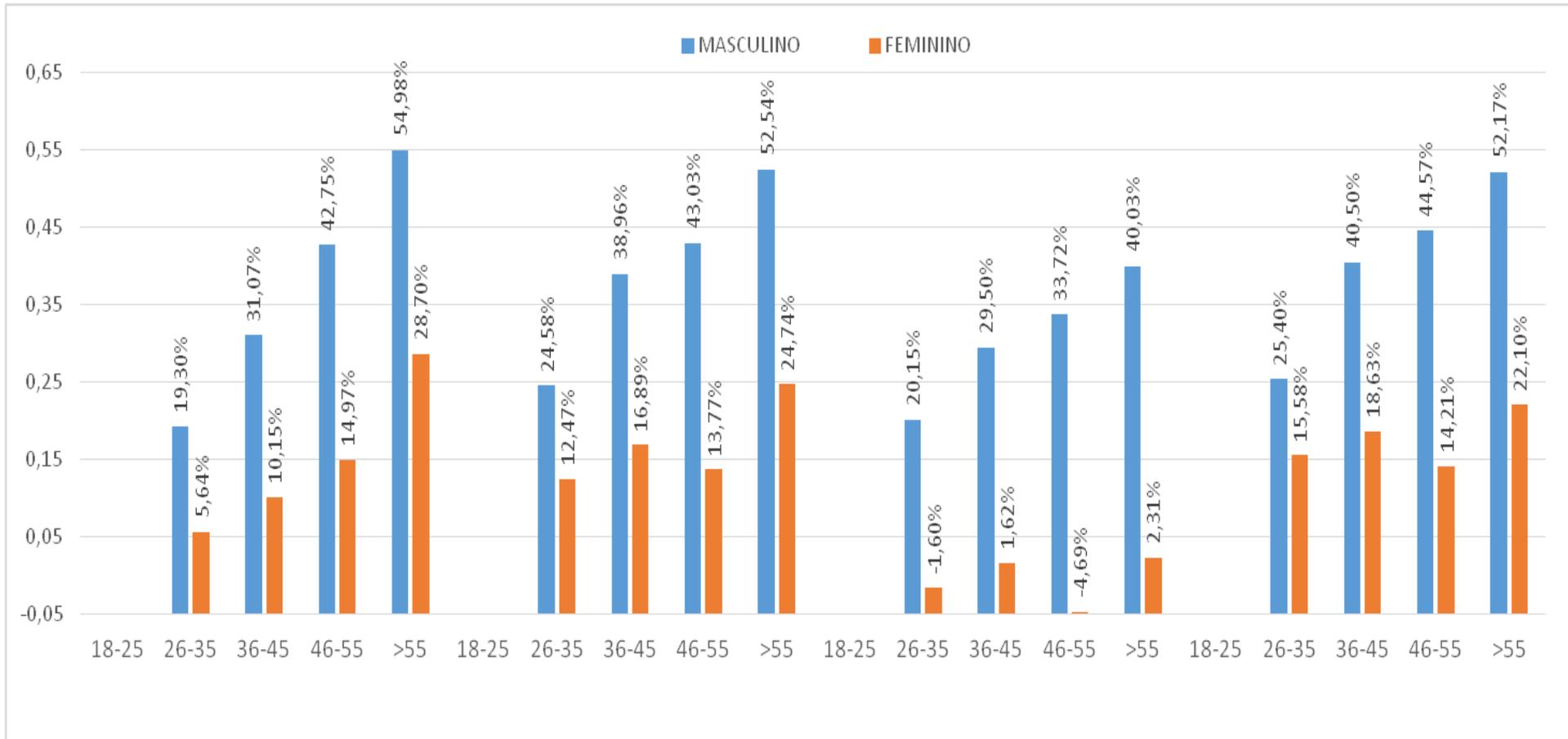


Gráfico 15. Análise por Amplitude
Fonte: Elaborado pelo Autor (2017).

4.3. RESULTADO DAS COTAÇÕES DE SEGURO

A demonstração dos resultados dos cálculos de seguros de automóveis efetuados nas 4 seguradoras, serão comparados aos gráficos do item 4.2 p.56 com o intuito de se verificar as semelhanças ou diferenças dos valores cobrados de seguro em relação aos sinistros registrados por esse perfil de condutor.

Foi possível observar na Tabela 13 que o condutor com idade de 60 anos pagará menor valor em todas as seguradoras em relação as demais idades calculadas e o condutor com 22 anos pagará o maior valor, o que corrobora com o gráfico 5 onde a ordem do risco permanece igual. A seguradora HDI em particular chegou a oferecer um seguro 62,66% mais barato para essa idade do que para uma pessoa de 22 anos, esse valor mostra ser discrepante quando relacionado ao Gráfico 5 onde verifica-se que o PREJUÍZO causado pelo condutor com mais de 55 anos é apenas 28,70 % menor que para faixa etária de 18 a 25 anos.

Tabela 13 - Cotação para o gênero feminino no Estado do RJ

CEP 28030-335 RJ	Idade					Amplitude percentual
	22 anos	31 anos	41 anos	51 anos	60 anos	
Seguradoras						
SULAMERICA	2326,80	1820,76	1754,23	1636,22	1513,13	34,97%
TOKIO	3255,71	2334,41	2051,15	1952,75	1917,99	41,09%
HDI	3440,19	1451,24	1344,00	1296,34	1284,42	62,66%
ALLIANZ	2822,12	2147,65	2076,49	2067,39	1980,36	29,83%

Fonte: Elaborado pelo Autor (2017)

Na Tabela 14 apenas na SULAMERICA o valor do seguro não foi decrescente acompanhando o crescimento da faixa etária de cálculo, essa oscilação corrobora com o que é observado no Figura 13 onde verifica-se que a faixa etária 3 ficou em 4º na lista de risco o que faria dela a segunda mais barata assim como apresentado. Mas em todas o perfil de 60 anos foi o mais barato e o de 22 anos o mais caro corroborando também com gráfico 11, sendo que a amplitude do Gráfico 11 atingiu o

valor máximo de 22,10 % apresentando discrepância de todos os resultados da Tabela 14.

Tabela 14 - Cotação para o gênero feminino no Estado de Minas Gerais

CEP 39900-000 MG	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	2525,92	1976,62	1769,55	1796,4	1651,03	34,64%
TOKIO	3042,77	2214,02	1948,71	1855,02	1812,51	40,43%
HDI	3409,92	2617,46	2424,69	2339,02	2317,6	32,03%
ALLIANZ	3361,16	2525,06	2443,46	2437,47	2341,15	30,35%

Fonte: Elaborado pelo Autor (2017)

Na Tabela 15 foi observado que as amplitudes percentuais variaram bastante de uma seguradora para outra, a seguradora ALLIANZ obteve uma amplitude percentual de 30,57% entre o maior e o menor valor de cotação e no gráfico 7 registrou-se 24,74% se tornando a mais próxima. Observa-se também que o valor da simulação é decrescente à medida que se aumenta a idade do condutor até atingir o menor valor de cotação no indivíduo com 60 anos, o que não converge com o demonstrado na Figura 9, onde a ordem de risco foi: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 46 a 55 anos” - “entre 36 a 45 anos” - “> que 55 anos”. 1 – 2 – 4 – 3 – 5.

Tabela 15 - Cotação para o gênero feminino no Estado de São Paulo

CEP 18133-445 SP	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	1698,57	1547,22	1513,54	1476,89	1430,11	15,81%
TOKIO	2954,83	2129,34	1873,94	1785,01	1752,26	40,70%
HDI	3917,58	2972,4	2742,5	2640,32	2614,77	33,26%
ALLIANZ	3950,37	3010,76	2901,82	2881,23	2742,84	30,57%

Fonte: Elaborado pelo Autor (2017)

Na Tabela 16 a seguradora SULAMERICA diferente das demais, ao realizar o cálculo do seguro mostrou que um condutor de 41 anos irá pagar um valor menor de seguro do que todas as outras idades utilizadas no cálculo, às demais seguradoras tiveram seus maiores valores no de 22 anos e foram reduzindo de acordo com que se aumentava a idade até atingir o mínimo com 60 anos. Foi possível observar também que na seguradora SULAMERICA os valores foram bem próximos em todos os cálculos atingindo uma amplitude máxima de apenas 13,06%, a ordem dos valores não convergiu em nenhuma cotação com o demonstrado no Gráfico 9 que seria: “entre 46 a 55 anos” – “entre 26 a 35 anos” - “entre 18 a 25 anos”– “entre 36 a 45 anos” — “> que 55 anos”.

Tabela 16 - Cotação para o gênero feminino no Estado do Espírito Santo

CEP 29075-910 ES	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	1631,24	1550,44	1497,42	1588,00	1545,75	8,20%
TOKIO	3019,91	2142,23	1881,51	1792,86	1772,47	41,31%
HDI	4153,54	3137,32	2890,15	2780,30	2752,83	33,72%
ALLIANZ	2800,12	2110,73	2042,86	2037,47	1957,01	30,11%

Fonte: Elaborado pelo Autor (2017)

Na Tabela 17 o valor da cotação de seguro nas 4 seguradoras se manteve decrescente na medida que se aumentou a idade do condutor, a cotação para uma pessoa de 60 anos chegou a ser 53,89% mais barata do que para uma de 22 anos (a seguradora HDI obteve um valor menor e mais distante das demais). É possível ver que a SULAMERICA possui maior semelhança nos valores da amplitude percentual com a amplitude do Gráfico 4, onde o percentual de PREJUÍZO foi 54,98% menor para faixa etária de maiores que 55 anos do que para faixa etária de 18 a 25 anos.

Tabela 17 - Cotação para o gênero masculino no Estado do Rio de Janeiro

CEP 28030-335 RJ	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	3517,84	2366,46	1880,00	1711,10	1622,00	53,89%
TOKIO	3904,00	2753,23	2373,28	2220,84	2140,92	45,16%
HDI	3678,5	2876,92	2681,95	2595,29	2573,63	30,04%
ALLIANZ	3962,76	2664,24	2276,33	2118,21	2042,4	48,46%

Fonte: Elaborado pelo Autor (2017)

Na Tabela 18 observa-se que nas 4 seguradoras as cotações que tiveram valores maiores foram com o perfil de 22 anos, valores esses que foram decrescendo até atingir o valor mínimo no perfil de 60 anos convergindo com a ordem de risco da Figura 12, a amplitude percentual da Tabela 18 atingiu o valor máximo de 57,44% e quando comparado com o Gráfico 10 é possível se observar uma convergência de informações onde nele mostra que uma pessoa com mais de 55 anos ofereceu em média 52,17% menos de PREJUÍZO do que uma pessoa na faixa etária de 18 a 25 anos.

Tabela 18 - Cotação para o gênero masculino no Estado de Minas Gerais

CEP 39900-000 MG	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	4056,05	2672,58	1894,8	1817,52	1726,29	57,44%
TOKIO	3718,28	2649,9	2284,59	2135,33	2046,9	44,95%
HDI	3808,67	2853,06	2660,29	2574,62	2553,2	32,96%
ALLIANZ	4693,79	3127,35	2674,28	2491,07	2407,7	48,70%

Fonte: Elaborado pelo Autor (2017)

A Tabela 19 a idade de 22 anos foi a que teve maior valor de cotação nas 4 seguradoras, os valores reduziram à medida que foi aumentando a idade do motorista convergindo com o Gráfico 6. A amplitude percentual desses valores atinge até 44,41% e a Figura 8 mostra que um segurado com mais de 55 anos

apresentou um PREJUÍZO 52,54% menor do que para umas pessoas de 18 a 25 anos.

Tabela 19 - Cotação para o gênero masculino no Estado de São Paulo

CEP 18133-445 SP	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	2488,52	1805,99	1548,51	1488,01	1470,70	40,90%
TOKIO	3549,37	2513,4	2169,4	2030,98	1973,03	44,41%
HDI	4198,59	3253,41	3023,49	2921,32	2895,77	31,03%
ALLIANZ	5660,41	3787,1	3205,19	2966,1	2873,36	49,24%

Fonte: Elaborado pelo Autor (2017)

Na Tabela 20 com exceção da SULAMERICA os valores mais baixos registrados para o seguro veicular foram para o condutor de 60 anos e em todas o seguro mais caro ficou para a idade de 22 anos. A diferença entre as amplitudes percentuais entre as seguradoras foi alta, indo de 13,06 a 48,47% e a amplitude percentual máxima vista no gráfico 8 foi de 40,03. A ordem de risco vista na Figura 10 foi: “entre 18 a 25 anos” – “entre 26 a 35 anos” – “entre 36 a 45 anos” – “entre 46 a 55 anos” – “> que 55 anos” corroborando com os resultados das cotações da TOKIO, HDI e ALLIANZ.

Tabela 20 - Cotação para o gênero masculino no Estado do Espírito Santo

CEP 29075-910 ES	Idade					Amplitude percentual
Seguradoras	22 anos	31 anos	41 anos	51 anos	60 anos	
SULAMERICA	1760,28	1715,33	1434,93	1571,51	1530,45	13,06%
TOKIO	3542,93	2480,71	2141,22	2008,25	1950,27	44,95%
HDI	4455,54	3439,41	3192,24	3082,39	3054,92	31,44%
ALLIANZ	3905,98	2610,62	2234,22	2082,46	2012,8	48,47%

Fonte: Elaborado pelo Autor (2017)

5. CONSIDERAÇÕES FINAIS

5.1. CONCLUSÕES

A utilização do algoritmo SimpleKmeans permitiu verificar que para todos os dados analisados dos 4 Estados da Região Sudeste do Brasil, tanto o perfil masculino quanto o feminino, obtiveram os seguintes atributos em comum: envolvimento em acidentes por falta de atenção, as colisões foram traseiras, não se teve registro de vítimas, ocorreram em pleno dia com céu claro, em retas, o veículo que dirigiam era um automóvel VW/Gol 1.0 do ano de 2008.

Sem dúvidas é uma quantidade grande de dados repetidos, o que leva à uma grande descoberta de grupo de perfil de risco, vale ressaltar que 33 % em média de dados foram excluídos da análise por terem sido listados como dados errados e/ou ignorados, porém, as análises prosseguiram normalmente.

Levando em consideração que o algoritmo apenas analisou dados quantitativos de acidentes e não valores de prejuízos por eles causados ou outros sinistros além de acidentes como nos dados da SUSEP e que ainda abrangeu todos os registros de acidentes em rodovias federais sendo eles com ou sem seguro, verificou-se que a faixa etária de maior risco foi a de 26 a 35 anos:

- Estado do RJ, gênero masculino: 31 anos;
- Estado do RJ, gênero feminino: 28 anos;
- Estado de SP, gênero masculino: 30 anos;

- Estado de SP, gênero feminino: 26 anos;
- Estado de MG, gênero masculino: 31 anos;
- Estado de MG, gênero feminino: 30 anos;
- Estado do ES, gênero masculino: 30 anos;
- Estado do ES, gênero feminino: 34 anos;

Analisando os dados da SUSEP foi possível concluir que:

- O maior GM foi registrado no Estado de MG, gênero masculino, faixa de 18 a 25 anos;
- O menor GM foi registrado no Estado de SP, gênero feminino, faixa > que 55 anos;
- Todos os valores de GM foram superiores para o gênero masculino em relação ao feminino em todos os registros;
- O maior TA foi registrado no Estado de SP, gênero feminino, faixa > que 55 anos;
- O menor TA foi registrado em MG, gênero masculino, faixa > que 55 anos;
- No Estado de SP e ES o gênero feminino registrou mais acidentes em sua proporção do que para os homens a partir da faixa etária de 36 a 45 anos e os Estados de RJ e MG também registraram esses dados porém a partir da faixa etária de 46 a 55 anos;
- O maior PREJUÍZO foi registrado no Estado do RJ, gênero masculino, faixa de 18 a 25 anos;
- O menor PREJUÍZO foi registrado no Estado de MG, gênero feminino, faixa > que 55 anos;
- Todos os valores de PREJUÍZO foram superiores para o gênero masculino em relação ao feminino em todos os registros;

Ao relacionar as figuras representativas criados a partir da base de dados da SUSEP com os resultados das cotações de seguro realizadas é possível se concluir

que as análises nem sempre corroboraram, pode-se dizer que na maioria a ordem de risco gerada pelas tabelas da SUSEP e os valores cobrados pelas seguradoras seguiram a mesma ordem, mas a discrepância dos valores cobrados entre as faixas etárias mostrou-se ser elevada. As cotações realizadas mostraram que atualmente no Brasil apenas uma variável observacional foi capaz de influenciar no cálculo e oferecer um seguro até 62,66% mais barato para uma pessoa de 60 anos do que para uma de 22 anos mostrando ainda que em todos Estados e seguradoras que foram estudados o condutor de 18 a 25 anos pagará maior valor do que qualquer outro de outra faixa etária.

Alguns países afirmam que 51% dos motoristas abaixo de 25 anos pouparia dinheiro sobre seus planos de seguro com o monitoramento de suas ações no trânsito e sendo oferecidos descontos pelo bom comportamento, como o sistema PAYD – Pay-As-You-Drive (Pague como você dirige), que é uma forma de se calcular o seguro, onde um aparelho é instalado no veículo e é possível obter descontos caso aja pouca e correta utilização do próprio, esse aparelho é capaz de medir a quilometragem média rodada, podendo gravar dados de condução, velocidades atingidas.

Para cada perfil de condutor e cada localidade onde foi realizado o cálculo do seguro obteve-se um tipo de análise diferente, o que mostra que apesar de ser possível se basear nos registros de sinistros passados para se calcular o prêmio do seguro, há uma forma que pode ser mais eficiente que seria agregar as análises comportamentais como comportamento individual no trânsito, comportamento de crédito no mercado, conduta em redes sociais, entre outros novos fatores que as empresas de seguro estudam para prever os riscos inerentes a cada perfil de condutor.

Os estudos mostram que apesar das pesquisas sobre seguros terem começado em 1927 e que desde então a preocupação dos autores é bem elevada, ainda há muito que se fazer e pesquisar, em sua grande maioria os autores comentava sobre insuficiência de acesso a dados tanto por parte de órgãos e seguradoras, quanto pelo conhecimento da veracidade das informações passadas pelo cliente na hora de uma contratação de seguro.

Sendo assim, se faz necessário um sistema mais eficiente que calcule a cotação de acordo com o risco individual oferecido por cada condutor, à prévia

classificação efetuada pelas seguradoras no Brasil pode estar ultrapassado e pode estar inibindo um grande número de condutores a contratar seguro logo quando se adquire o primeiro veículo por efeito do alto valor oferecido.

5.2. SUGESTÕES DE TRABALHOS FUTUROS

Para buscar se aprofundar mais ainda no assunto, sugere-se que o ponto de partida da revisão bibliográfica seja buscado em outras bases além da SCOPUS, com o intuito de descobrir novos estudos na área.

Como trabalho futuro, podem-se realizar pesquisas que possam comprovar a eficiência do uso da variável comportamental na análise de riscos, tornando os cálculos mais pessoais, o que reduziria os valores para perfis de menos riscos e compensaria nos maiores.

O fato de uma possível relação entre o número de acidentes e o número de segurados deve ser melhor estudada (fatores da seleção adversa e seleção propícia), a ligação desses fatores transformaria essa pesquisa numa questão de saúde pública.

Após o registro de um carro em qualquer país da União Europeia é obrigatório a contratação do seguro contra terceiros que irá cobrir apenas os danos causados ao outro veículo, sendo facultativo a contratação do seguro completo que inclui as demais coberturas. Sugere-se como estudo futuro uma verificação da viabilidade e o levantamento dos prós e contras de se ter essa obrigatoriedade no Brasil.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ADAMOS, G.; NATHANAIL, E. How to train safe drivers: Setting up and evaluating a fatigue training program. **Transport and Telecommunication**, v. 16, n. 1, p. 9–20, 2015. Disponível em: <<https://www.degruyter.com/view/j/ttj.2015.16.issue-1/ttj-2015-0002/ttj-2015-0002.xml>>. Acesso em: 20 jan. 2017.

DAMS, D. O. Fisco, inc., the equity-funding of automobile insurance. **EDPACS**, v. 5, n. 6, p. 1–5, 1977. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/07366987709449629?tab=permissions&scroll=top>>. Acesso em: 20 jan. 2017.

APTE, C. et al. Business applications of data mining. **Communications of the ACM**, v. 45, n. 8, p. 49–53, 2002. Disponível em: <<http://dl.acm.org/citation.cfm?id=545178>>. Acesso em: 20 jan. 2017.

ARBIS, D; DIXIT, V. V; RASHIDI, T. H. Impact of risk attitudes and perception on game theoretic driving interactions and safety. **Accident Analysis and Prevention**, v. 94, p. 135–142, 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/27289391>>. Acesso em: 20 jan. 2017.

ATKINSON, M. M.; NIGOL, R. A. Selecting policy instruments: Neo-institutional and rational choice interpretations of automobile insurance in ontario. **Canadian Journal of Political Science**, v. 22, n. 1, p. 107–135, 1989. Disponível em: <<https://www.cambridge.org/core/journals/canadian-journal-of-political-science-revue-canadienne-de-science-politique/article/selecting-policy-instruments-neoinstitutional-and-rational-choice-interpretations-of-automobile-insurance-in-ontario/F7B87DF99E5E6B2E3E1513385D71A947#>>. Acesso em: 20 jan. 2017

BAILEY, R. A.; SIMON, L. J. Two Studies in Automobile Insurance Ratemaking. **ASTIN Bulletin**, v. 1, n. 4, p. 192–217, 1960. Disponível em: <<http://www.actuaries.org/LIBRARY/ASTIN/vol1no4/192.pdf>>. Acesso em: 20 jan. 2017.

BEGG, D; BROOKLAND, R. R. Participation in driver education/training courses during graduated driver licensing, and the effect of a time-discount on subsequent traffic offenses: Findings from the New Zealand Drivers Study. **Journal of Safety Research**, v. 55, p. 13–20, 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022437515000523>>. Acesso em: 20 jan. 2017.

BEIRLANT, J. et al. Statistical risk evaluation applied to (Belgian) car insurance. **Insurance Mathematics and Economics**, v. 10, n. 4, p. 289–302, 1992. Disponível em: <<https://lirias.kuleuven.be/handle/123456789/38642>>. Acesso em: 20 jan. 2017.

BERNARD, D. Is the risk of product market predation a cost of disclosure? **Journal of Accounting and Economics**, v. 62, n. 2–3, p. 305–325, 2016. Disponível em: <<http://www.simon.rochester.edu/faculty-and-research/conferences/journal-of-accounting-and-economic-conference/papers/download.aspx?id=12802>>. Acesso em: 20 jan. 2017.

BLOCK, W. Private roads, competition, automobile insurance and price controls. **Competitiveness Review**, v. 8, n. 1, p. 55–64, 1998. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/eb046362?src=recsys>>. Acesso em: 20 jan. 2017.

BLOWS, S. et al. Car insurance and the risk of car crash injury. **Accident Analysis and Prevention**, v. 35, n. 6, p. 987–990, 2003. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/12971933>>. Acesso em: 20 jan. 2017.

BOJ, E. et al. Implementing PLS for distance-based regression: **Computational issues: Computational Statistics**, v. 22, n. 2, p. 237–248, 2007. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/12971933>>. Acesso em: 20 jan. 2017.

BOLETIN DE INFORMACION - ILUSTRE COLEGIO OFICIAL DE ODONTOLOGOS Y ESTOMATOLOGOS 2. A REGION (CATALUNA). **Automobile insurance**, v. 9, n. 45, p. 22–23, 1965.

BURLACU, E. Risk methodology in vehicle insurance. **Quality - Access to Success**, v. 13, n. SUPPL.3, p. 300–304, 2012. Disponível em: <<http://web.a.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authType=crawler&jrnl=15822559&AN=78234618&h=X2kymA%2bxhB2XVSduOACUh9ef6wTuRjdbdw4BTDEYbl%2fVarncTMYwDaseGFTPQ%2fwNUjl3CF3sOId9qzbixCssCg%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authType%3dcrawler%26jrnl%3d15822559%26AN%3d78234618>>. Acesso em: 20 jan. 2017.

CANNON, E.; CIPRIANI, G. P.; BAZAR-ROSEN, K. More for less? Puzzling selection effects in the insurance market. **Oxford Economic Papers**, v. 68, n. 4, p. 879–897, 2016. Disponível em: <<https://academic.oup.com/oep/article-abstract/68/4/879/2503389/More-for-less-Puzzling-selection-effects-in-the?redirectedFrom=fulltext>>. Acesso em: 20 jan. 2017.

CHAPPELL, D.; NORMAN, J. M. Optimal, near-optimal and rule-of-thumb claiming rules for a protected bonus vehicle insurance policy. **European Journal of Operational Research**, v. 41, n. 2, p. 151–156, 1989. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0377221789903779>>. Acesso em: 20 jan. 2017.

CONFEDERAÇÃO NACIONAL DAS EMPRESAS DE SEGUROS GERAIS, PREVIDÊNCIA PRIVADA E VIDA, SAÚDE SUPLEMENTAR E CAPITALIZAÇÃO (CNSEG). **Relatório de Sustentabilidade do Setor de Seguros**: 2016. Disponível em: <<http://www.cnseg.org.br/lumis/portal/file/fileDownload.jsp?fileId=8A8AA88A561E2ED60156330136B51F8A>>. Acesso em: 20 jan. 2017.

COSTA, H. G. Modelo para webibliomining: proposta e caso de aplicação. **Revista FAE**, v. 13, n. 1, p. 115–126, 2010. Disponível em: <<https://www.yumpu.com/pt/document/view/34551757/modelo-para-webibliomining-proposta-e-caso-de-aplicacao>>. Acesso em: 20 jan. 2017.

CUMMINS, J. D.; PHILLIPS, R. D.; WEISS, M. A. The incentive effects of no-fault automobile insurance. **Journal of Law and Economics**, v. 44, n. 2, p. 427–464, 2001. Disponível em: <http://www.jstor.org/stable/10.1086/322818?seq=1#page_scan_tab_contents>. Acesso em: 20 jan. 2017.

DEVLIN, R. A. Some welfare implications of no-fault automobile insurance. **International Review of Law and Economics**, v. 10, n. 2, p. 193–205, 1990. Disponível em: <http://econpapers.repec.org/article/eeeirlaec/v_3a10_3ay_3a1990_3ai_3a2_3ap_3a193-205.htm>. Acesso em: 20 jan. 2017.

DIJKSTERHUIS, C. et al. In-car usage-based insurance feedback strategies. A comparative driving simulator study. **Ergonomics**, v. 59, n. 9, p. 1158–1170, 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/26653393>>. Acesso em: 20 jan. 2017.

_____; et al. The impact of immediate or delayed feedback on driving behaviour in a simulated Pay-As-You-Drive system. **Accident Analysis and Prevention**, v. 75, p. 93–104, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/25460096>>. Acesso em: 20 jan. 2017.

ELLISON, A.B., GREAVES, S.P., BLIEMER, M. C. J. Driver behaviour profiles for road safety analysis. **Accident Analysis and Prevention**, v. 76, p. 118–132, 2015. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25638284>>. Acesso em: 20 jan. 2017.

_____;_____;_____. Evaluating changes in driver behaviour: A risk profiling approach. **Accident Analysis and Prevention**, v. 75, p. 298–309, 2014. Disponível em: <https://www.researchgate.net/publication/270007418_Evaluating_changes_in_driver_behaviour_A_risk_profiling_approach>. Acesso em: 20 jan. 2017.

EUROPA.EU. **Validade do seguro automóvel na EU**: Validade do seguro obrigatório e do seguro facultativo Disponível em: <http://europa.eu/youreurope/citizens/vehicles/insurance/validity/index_pt.htm>. Acesso em: 14 fev. 2017.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **Advances in Knowledge Discovery and Data Mining**. Menlo Park, EUA: AAAI Press, p. 611, 1996. Disponível em: <<http://dl.acm.org/citation.cfm?id=257938>>. Acesso em: 14 fev. 2017.

GOLDEN, L. L. et al. Empirical Evidence on the Use of Credit Scoring for Predicting Insurance Losses with Psycho-social and Biochemical Explanations. **North American Actuarial Journal**, v. 20, n. 3, p. 233–251, 2016. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/10920277.2016.1209118>>. Acesso em: 14 fev. 2017.

GÓMEZ-DÉNIZ, E. Bivariate credibility bonus-malus premiums distinguishing between two types of claims. **Insurance: Mathematics and Economics**, v. 70, p. 117–124, 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167668716300750>>. Acesso em: 14 fev. 2017.

HARRINGTON, S. E.; NIEHAUS, G. Race, redlining, and automobile insurance prices. **Journal of Business**, v. 71, n. 3, p. 439–469, 1998. Disponível em: <http://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1000&context=fin_facpu>. Acesso em: 14 fev. 2017.

HAUGEN, M.; MOGER, T. A. Frailty modelling of time-to-lapse of single policies for customers holding multiple car contracts. **Scandinavian Actuarial Journal**, v. 2016, n. 6, p. 489–501, 2016. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/03461238.2014.963886?journalCode=sact20>>. Acesso em: 14 fev. 2017.

HEMENWAY, D. Propitious selection in insurance. **Journal of Risk and Uncertainty**, v. 5, n. 3, p. 247–251, 1992. Disponível em: <<https://link.springer.com/article/10.1007/BF00057881>>. Acesso em: 14 fev. 2017.

HSU, Y.-C.; CHOU, P.-L.; SHIU, Y.-M. An examination of the relationship between vehicle insurance purchase and the frequency of accidents. **Asia Pacific Management Review**, v. 21, n. 4, p. 231–238, 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1029313216301762>>. Acesso em: 14 fev. 2017.

JAIN, A. Data Clustering: 50 Years Beyond K-Means. **Pattern Recognition Letters**, Michigan - USA, v. 31 p. 651–666 2008 (2010). Disponível em: <http://www.ppgia.pucpr.br/~fabricio/ftp/Roges/JainClustering_PRL10.pdf>. Acesso em: 14 fev. 2017.

JOLOWICZ, J. A. Vicarious liability-motor vehicle-insurance. **The Cambridge Law Journal**, v. 29, n. 2, p. 195–198, 1971. Disponível em: <<https://www.cambridge.org/core/journals/cambridge-law-journal/article/vicarious-liabilitymotor-vehicleinsurance/54D9C36DC2DE09AB087B2B8D9762D2B4>>. Acesso em: 14 fev. 2017.

JUNG, J. On automobile insurance ratemaking. **ASTIN Bulletin**, v. 5, n. 1, p. 41–48, 1968. Disponível em: <<https://www.casact.org/library/astin/vol5no1/41.pdf>>. Acesso em: 14 fev. 2017.

LEDO, Bruno Cesar Aurichio. Competição em preços entre corretores de seguros de automóveis. **Estudos Economicos**, São Paulo, v. 41, n. 4, p. 719–741, 2011. Disponível em: <<http://www.revistas.usp.br/ee/article/view/36061/38780>>. Acesso em: 14 fev. 2017.

LEE, Y.-C.; WINSTON, F. K. Stress induction techniques in a driving simulator and reactions from newly licensed drivers. **Transportation Research Part F: Traffic Psychology and Behaviour**, v. 42, p. 44–55, 2016. Disponível em: <https://www.researchgate.net/publication/308129113_Stress_induction_techniques_in_a_driving_simulator_and_reactions_from_newly_licensed_drivers>. Acesso em: 14 fev. 2017.

LI, C.-S.; LIU, C.-C. On compulsory per-claim deductibles in automobile insurance. **Geneva Papers on Risk and Insurance Theory**, v. 28, n. 1, p. 25–32, 2003. Disponível em: <<https://link.springer.com/article/10.1023/A:1022191714312>>. Acesso em: 14 fev. 2017.

LOMAN, H. J. Automobile Insurance. **The Annals of the American Academy of Political and Social Science**, v. 161, n. 1, p. 85–90, 1932. Disponível em: <<https://www.safetylit.org/week/journalpage.php?jid=5737>>. Acesso em: 14 fev. 2017.

_____. Compulsory Automobile Insurance. **The Annals of the American Academy of Political and Social Science**, v. 130, n. 1, p. 163–170, 1927. Disponível em: <<https://www.safetylit.org/week/journalpage.php?jid=5737>>. Acesso em: 14 fev. 2017.

MANDIC, D. Multi-criteria AHP analysis in risk assessment for motor vehicle insurance. **Metalurgia International**, v. 18, n. 1, p. 128–131, 2013. Disponível em: <https://www.researchgate.net/publication/289170138_Multi-criteria_AHP_analysis_in_risk_assessment_for_motor_vehicle_insurance>. Acesso em: 14 fev. 2017.

MARONEY, P. F.; VICKORY, F. A. Discrimination in automobile insurance: issues and remedies. **American Business Law Journal**, v. 24, n. 2, p. 269–291, 1986.

MARTIN, D. B. Automobile Insurance: Canadian Accident-Free Classification System. **ASTIN Bulletin**, v. 1, n. 3, p. 123–133, 1960. Disponível em: <<http://www.actuaries.org/LIBRARY/ASTIN/vol1no3/123.pdf>>. Acesso em: 14 fev. 2017.

MENDONÇA, A. P. **Seguros**: não é tudo a mesma coisa. Disponível em: <<http://www.legiscor.com.br/Artigos/Artigos/Seguros-nao-e-tudo-a-mesma-coisa.html>>. Acesso em: 16 fev. 2017.

MINUTOS SEGUROS. **Por que fazer um seguro auto?** Disponível em: <<https://www.minutoseguros.com.br/blog/carro/por-que-fazer-seguro-auto>>. Acesso em: 5 mar. 2017.

MUNDEN, J. M. Some Analyses of Car Insurance Claim-Rates. **ASTIN Bulletin**, v. 2, n. 2, p. 182–197, 1962. Disponível em: <<https://www.cambridge.org/core/services/aop-cambridge-core/content/view/A448CBF3F2D93C8D3D3231DAE614220A/S0515036100009946a.pdf/some>>. Acesso em: 16 fev. 2017.

NANNI, M. et al. Driving profiles computation and monitoring for car insurance CRM. **ACM Transactions on Intelligent Systems and Technology**, v. 8, n. 1, 2016. Disponível em: <<http://dl.acm.org/citation.cfm?id=2912148>>. Acesso em: 16 fev. 2017.

PÉREZ-FRUCTUOSO, M. J.; PÉREZ, A. G. Analyzing solvency with extreme value theory: An application to the Spanish motor liability insurance market . **Innovar**, v. 20, n. 36, p. 35–48, 2010. Disponível em: <<http://www.scielo.org.co/pdf/inno/v20n36/20n36a04.pdf>>. Acesso em: 16 fev. 2017.

POLÍCIA RODOVIÁRIA FEDERAL (PRF). **Dicionário de Variáveis**: infrações: 2016. Brasília: PRF, 2017. Disponível em: <https://www.prf.gov.br/portal/dados-abertos/dicionario-de-dados-infracoes/at_download/file>. Acesso em: 20 fev 2017.

_____. **DPRF**: serviço de Nada Consta informa se o veículo possui multas por infringir o Código de Trânsito Brasileiro (CTB) extraídas pela Polícia Rodoviária Federal (PRF).. Disponível em: <<https://www.prf.gov.br/portal/dados-abertos>>. Acesso em: 6 dez. 2016.

QUALIDADETOTAL. **Diagrama de Pareto**. Disponível em: <<http://apostilasdaqualidade.com.br/ferramentas/diagrama-de-pareto/>>. Acesso em: 5 mar. 2017.

RIBEIRO, A H. P.; GRISI, C. C. H. Marketing de relacionamento como fator-chave de sucesso no mercado de seguros. **Revista de Administração de Empresas (RAE)**, v. 39, n. 1, p. 31–41, 1999. Disponível em: <<http://www.scielo.br/pdf/rae/v39n1/v39n1a05.pdf>>. Acesso em: 6 dez. 2016.

ROSS, V. et al. Investigating risky, distracting, and protective peer passenger effects in a dual process framework. **Accident Analysis and Prevention**, v. 93, p. 217–225, 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0001457516301610>>. Acesso em: 6 dez. 2016.

SABITHA, R.; KARTHIK, S. Convalescing Cluster Configuration Using a Superlative **The Scientific World Journal**, v. 2015, 2015. Disponível em: <<https://www.hindawi.com/journals/tswj/2015/180749/>>. Acesso em: 6 dez. 2016.

SARABIA, J. M. et al. Risk aggregation in multivariate dependent Pareto distributions. **Insurance: Mathematics and Economics**, v. 71, p. 154–163, 2016. Disponível em: <http://econpapers.repec.org/article/eeeinsuma/v_3a71_3ay_3a2016_3ai_3ac_3ap_3a154-163.htm>. Acesso em: 6 dez. 2016.

SCOPUS: ELSEVIER BASES DE DADOS. Disponível em: <<http://www.americalatina.elsevier.com/corporate/scopus.php>>. Acesso em: 9 jan. 2017.

SEGOVIA, González, M. M.; CONTRERAS Rubio, I.; MAR Molinero, C. Efficiency assessment in the automobile insurance sector . **Revista de Metodos Cuantitativos para la Economía y la Empresa**, v. 7, p. 57–76, 2009. Disponível em: <https://www.researchgate.net/publication/288414672_Efficiency_assessment_in_the_automobile_insurance_sector>. Acesso em: 9 jan. 2017.

_____.; GUERRERO CASAS, F. M.; HERRANZ PEINADO, C. P. Premium allocation in the car insurance by using functional principal component analysis . **Revista de Metodos Cuantitativos para la Economía y la Empresa**, v. 4, p. 56–74, 2007. Disponível em: <https://www.researchgate.net/publication/26490935_Asignacion_de_primas_en_el_seguro_del_automovil_utilizando_el_Analisis_en_ComponentesPrincipales_Funcionales_Premium_allocation_in_the_car_insurance_by_using_Functional_Principal_Component_Analysis>. Acesso em: 9 jan. 2017.

SEGURO AUTOMÓVEL. **Cálculo-valor-seguro**. Disponível em: <<http://www.segurosautomovel.com.br/perguntas-frequentes/como-e-calculado-o-valor-do-seguro>>. Acesso em: 9 jan. 2017.

SHI, P.; ZHANG, W. A Test of Asymmetric Learning in Competitive Insurance With Partial Information Sharing. **Journal of Risk and Insurance**, v. 83, n. 3, p. 557–578, 2016. Disponível em: <<http://www.segurosautomovel.com.br/perguntas-frequentes/como-e-calculado-o-valor-do-seguro>>. Acesso em: 9 jan. 2017.

SILWATTANANUSARN, TUAMSUK, Kulthida . Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012. **International Journal of Data Mining and Knowledge Management Process (IJDMP)**, v. 2, n. 5, 2012. Disponível em: <https://www.researchgate.net/publication/232143760_Data_Mining_and_Its_Applications_for_Knowledge_Management_A_LiteratureReview_from_2007_to_2012>. Acesso em: 9 jan. 2017.

STELLWAGEN, H. P. Automobile Insurance. **The Annals of the American Academy of Political and Social Science**, v. 130, n. 1, p. 154–162, 1927. Disponível em: <https://www.researchgate.net/publication/232143760_Data_Mining_and_Its_Applications_for_Knowledge_Management_A_LiteratureReview_from_2007_to_2012>. Acesso em: 9 jan. 2017.

STERN, R. A. et al. Office-Based Assessment of At-Risk Driving in Older Adults with and Without Cognitive Impairment. **Journal of Geriatric Psychiatry and Neurology**, v. 29, n. 6, p. 352–360, 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/27647790>>. Acesso em: 9 jan. 2017.

SUPERINTENDÊNCIA DE SEGUROS PRIVADOS (SUSEP). **AUTOSEG**: sistema de estatísticas de automóveis da SUSEP. Disponível em: <<http://www2.susep.gov.br/menuestatistica/Autoseg/menu1.aspx>>. Acesso em: 1 dez. 2016.

THOMAS, D. J.; YOUNG, P. Assuring the Ability to Pay: A Comparative Look at Automobile Insurance in Minnesota and Manitoba. **American Review of Canadian Studies**, v. 23, n. 2, p. 203–216, 1993.

TRIBUNAL DE CONTAS DA UNIÃO (TCU). Relatório de gestão do exercício de 2015. Disponível em: <<http://portal.tcu.gov.br/contas/contas-e-relatorios-de-gestao/contas-do-exercicio-de-2015.htm>>. Acesso em: 20 fev 2017..

VATTER, C. Simulation of automobile insurance bonus-malus systems. **American Journal of Mathematical and Management Sciences**, v. 17, n. 3–4, p. 329–357, 1997. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/01966324.1997.10737443?needAccess=true&journalCode=umms20>>. Acesso em: 20 fev 2017..

WEIDNER, W.; TRANSCHEL, F. W. G.; WEIDNER, R. Classification of scale-sensitive telematic observables for risk individual pricing. **European Actuarial Journal**, v. 6, n. 1, p. 3–24, 2016. Disponível em: <<https://link.springer.com/article/10.1007/s13385-016-0127-x>>. Acesso em: 20 fev 2017..

WEKA : UNIVERSITY WAIKATO. **Weka 3**: Data Mining Software in Java. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 21 fev. 2017.

YANG, L.; DENG, M. Based on k-Means and Fuzzy k-Means Algorithm Classification of Precipitation In: INTERNATIONAL SYMPOSIUM ON COMPUTATIONAL INTELLIGENCE AND DESIGN, 3, 2010, 29-31 oct; Hangzhou, China. **Proceedings...** New York-USA: IEEE, 2011. p. 218–221. Disponível em: <<http://ieeexplore.ieee.org/document/5692703/>>. Acesso em: 21 jan. 2017.

YU, J.-M.; HAO, X.-D. An empirical rating actuarial model with claim size following lognormal distribution in automobile insurance. **Shanghai Jiaotong Daxue Xuebao/Journal of Shanghai Jiaotong University**, v. 42, n. 11, p. 1836–1838, 2008. Disponível em: <https://www.researchgate.net/publication/288450336_An_empirical_rating_actuarial_model_with_claim_size_following_lognormal_distribution_in_automobile_insurance>. Acesso em: 21 jan. 2017.

YUZVOVICH, L. et al. Vehicle insurance financial mechanism. **Review of European Studies**, v. 7, n. 3, p. 99–112, 2015. Disponível em: <<http://www.ccsenet.org/journal/index.php/res/article/view/45893/24836>>. Acesso em: 21 jan. 2017.

ZHAO, X.; ZHOU, X. Copula-based dependence between frequency and class in car insurance with excess zeros. **Operations Research Letters**, v. 42, n. 4, p. 273–277, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016763771400042X>>. Acesso em: 21 jan. 2017.