

UNESP
Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Engenharia
Programa de Pós Graduação em Engenharia de Produção

PRISCILA AMORIM DA COSTA

**ESTUDO DA SINISTRALIDADE NO MERCADO SECURITÁRIO DE VEÍCULOS:
UMA ABORDAGEM MULTIVARIADA**

BAURU-SP
2014

PRISCILA AMORIM DA COSTA

**ESTUDO DA SINISTRALIDADE NO MERCADO SECURITÁRIO DE VEÍCULOS:
UMA ABORDAGEM MULTIVARIADA**

Dissertação apresentada ao Departamento de Engenharia de Produção da Faculdade de Engenharia da UNESP de Bauru/SP, como requisito para obtenção do grau de Mestre em Engenharia de Produção. Orientador: Prof. Dr. Manoel Henrique Salgado

BAURU-SP
2014

Costa, Priscila Amorim.
Estudo da Sinistralidade no Mercado
Securitário de Veículos: Uma Abordagem
Multivariada / Priscila Amorim da Costa, 2014
87 f.

Orientador: Manoel Henrique Salgado

Dissertação (Mestrado)-Universidade Estadual
Paulista. Faculdade de Engenharia, Bauru, 2014


1. Sinistralidade. 2. Seguro de automóvel. 3.
Análise discriminante. I. Universidade Estadual
Paulista. Faculdade de Engenharia. II. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE Mestrado DE PRISCILA AMORIM DA COSTA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO, DO(A) FACULDADE DE ENGENHARIA DE BAURU.

Aos 10 dias do mês de novembro do ano de 2014, às 09:00 horas, no(a) Sala de Reuniões da Seção Técnica de Pós-graduação da Faculdade de Engenharia de Bauru, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. MANOEL HENRIQUE SALGADO do(a) Departamento de Engenharia de Produção / Faculdade de Engenharia de Bauru, Profa. Dra. GLADYS DOROTEA CACSIRE BARRIGA do(a) Departamento de Engenharia de Produção / Faculdade de Engenharia de Bauru, Prof. Dr. JOAO PEDRO ALBINO do(a) Departamento de Computação / Faculdade de Ciências de Bauru, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE Mestrado de PRISCILA AMORIM DA COSTA, intitulada "APLICAÇÃO DA ANÁLISE DISCRIMINANTE NO ESTUDO DA SINISTRALIDADE E SUAS RELAÇÕES NO MERCADO SECURITÁRIO DE VEÍCULOS - ESTUDO DE CASO EM UMA CORRETORA DE BAURU". Após a exposição, a discente foi arguida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADA. Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.


Prof. Dr. MANOEL HENRIQUE SALGADO


Profa. Dra. GLADYS DOROTEA CACSIRE BARRIGA


Prof. Dr. JOAO PEDRO ALBINO



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Bauru

PROPOSTA DE ALTERAÇÃO DO TÍTULO

A BANCA EXAMINADORA PROPÕE A ALTERAÇÃO DO TÍTULO DO TRABALHO DA ALUNA:
PRISCILA AMORIM DA COSTA

DE: "APLICAÇÃO DA ANÁLISE DISCRIMINANTE NO ESTUDO DA SINISTRALIDADE E SUAS
RELAÇÕES NO MERCADO SECURITÁRIO DE VEÍCULOS - ESTUDO DE CASO EM UMA
CORRETORA DE BAURU"

PARA:

ESTUDO DA SINISTRALIDADE NO MERCADO SECURITÁRIO
DE VEÍCULOS: UMA ABORDAGEM MULTIVARIADA.

Bauru, 10 de novembro de 2014.


Prof. Dr. Manoel Henrique Salgado
Orientador

DEDICATÓRIA

À Deus, que me tem sustentado e capacitado a cada dia.
Agradeço também a meus pais Roberto e Elisabete, à
minha avó querida Laudemira, às minhas irmãs Patrícia e
Paola, amigos e a todo esforço/APRENDIZADO
outorgado a mim pelo professor Manoel Henrique
Salgado.

AGRADECIMENTOS

Ao Prof.º Dr. Manoel Henrique pela orientação, amizade, ensinamentos e confiança depositada, que contribuíram para o meu crescimento.

A todos os meus amigos e familiares que, de alguma forma, torceram por mim e me ajudaram nesta conquista.

Aos professores e funcionários do Departamento de Engenharia de Produção da UNESP de Bauru/SP.

A todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

“Para realizar grandes conquistas, devemos não apenas agir, mas também sonhar; não apenas planejar, mas também acreditar.”

Anatole France

RESUMO

A seguradora que possui um controle eficaz dos riscos oferecidos por parte de seus segurados pode evitar transtornos do tipo falência e perda de lucratividade. O objetivo da pesquisa foi identificar as variáveis que definem o risco associado aos futuros compradores, possibilitando classificá-los em um de dois grupos, o sinistrado e o não sinistrado, em função de probabilidades definidas por um modelo multivariado. Para tanto, foram consideradas questões (variáveis) existentes no questionário de avaliação de risco das seguradoras e outras eleitas por corretores experientes. Para alcançar o objetivo, utilizou-se na análise a técnica estatística multivariada, conhecida como Análise Discriminante, com a finalidade de segregar os indivíduos em um dos dois grupos. Uma função discriminante foi construída a partir das variáveis independentes, associadas aos riscos, e da variável dependente, que abrange os dois grupos. Outros resultados como a estimação da regra de classificação, avaliação da qualidade de ajuste da regra de discriminação, estimação da probabilidade global de acertos e os testes relacionados com os pressupostos da análise discriminante foram apresentados. A amostra consultada foi composta por 2.000 segurados atendidos pela corretora, divididos em dois grupos: o primeiro composto por indivíduos sem sinistro; e o segundo com os que possuem uma ou mais ocorrências. O modelo possibilitou classificar os indivíduos pelos dois grupos, sendo que na amostra de desenvolvimento e de teste a classificação representou 69% de acerto. A separação se fez através de variáveis com graus maiores de importância. Foram eleitas: potência do carro, tempo de seguro, bônus tempo, uso de estacionamento e atividades profissionais definidas pelos cargos ocupados pelos clientes. Tais variáveis podem ser consideradas altamente discriminadoras, com base no coeficiente de contribuição para a discriminação. Este trabalho apresentou resultados que contrapõem o senso comum vigente no mercado segurador, no qual os técnicos têm por prática afirmar que o sexo e a idade determinam se um indivíduo terá menor ou maior quantidade de sinistros. Esperava-se que os coeficientes dessas duas variáveis fossem mais relevantes para o modelo, visto que o maior percentual de sinistros aponta para o sexo masculino.

Palavras chave: Sinistralidade, seguro de automóvel, análise discriminante.

ABSTRACT

The insurer that has an effective control of the risks involved by the insureds can avoid issues such as bankruptcy and loss of profitability. The purpose of the research was to identify the varieties that define the risk associated with future buyers, enabling classify them into one of two groups, the claimed and the unclaimed, based in probabilities defined by a multivariate model. Therefore, were questions considered (variables) existing on the risk evaluation of the insurers questionnaire and others nominated by the expert brokers of the insurers. To achieve the objective, was used in the analyses, the multivariate statistical technique, known as Discriminant Analyses, in order to segregate the individual into one of two groups. A discriminant function was constructed from the independent variables, associated to the risks, and from the dependent variable, which covers the two groups. Other results as estimation of the classification rule, evaluation of the quality of the discrimination rule settings, estimation of the overall probability of correct answers and tests related to the assumptions of discriminant analysis were presented. The studied sample was consisted of 2,000 insured served by the broker, divided into two groups: the first composed of individuals without claims; and the second with those who has one or more claims. The model enabled to classify individuals by the two groups, wherein in the development sample and test the classification represented 69% accuracy. The separation was carried out by variables with higher importance degrees. Were elected: car power, time of insurance, bonus, parking lot use, and professional activities defined by the insured occupation. Such variables can be considered highly discriminating, based on the contribution coefficient for the discrimination. This work has presented results that contradict the common sense in the insurance market, in which technicians say that sex and age determine whether an individual will have more or less quantity of claims. It was expected that the coefficients of these two variables were more relevant to the model, since the highest percentage of claims points to the males.

Keywords: Accidents, auto insurance, discriminant analysis.

LISTA DE FIGURAS

Figura 01 Estrutura da dissertação.....	18
Figura 02 Populações normais com mesma variabilidade e uma variável discriminante.....	37
Figura 03 Função discriminante de Fisher.....	39
Figura 04 Análise Discriminante.....	53

LISTA DE TABELAS

Tabela 01: Termos de Seguros.....	21
Tabela 02: Número esperado de classificados por caso.....	46
Tabela 03: <i>Dummys</i> para sexo.....	53
Tabela 04: <i>Dummys</i> para estado civil.....	54
Tabela 05: Resumo das potenciais variáveis para as análises.....	56
Tabela 06: Indicadores de variáveis quantitativas.....	58
Tabela 07: Distribuição dos segurados por grupo em relação ao sexo.....	59
Tabela 08: Distribuição dos segurados por grupo segundo a classe de idades.....	60
Tabela 09: Distribuição dos segurados por grupo quanto ao estado civil.....	60
Tabela 10: Distribuição dos segurados por grupo em relação a região.....	61
Tabela 11: Distribuição dos segurados por grupo em relação aos cargos.....	62
Tabela 12: Distribuição dos segurados por grupo em relação as classes de potência (cc).....	63
Tabela 13: Distribuição dos segurados por grupo em relação a idade do veículo.....	63
Tabela 14: Distribuição dos segurados por grupo em relação ao estacionamento.....	64
Tabela 15: Distribuição dos segurados por grupo em relação ao mês de contratação.....	65
Tabela 16: Distribuição dos segurados por grupo em relação às seguradoras.....	66
Tabela 17: Distribuição dos segurados por grupo em relação ao tempo de seguro.....	66
Tabela 18: Distribuição dos segurados por grupo em relação ao valor do seguro.....	67
Tabela 19: Dados da amostra subdividida.....	68

Tabela 20: Variáveis consideradas.....	69
Tabela 21: Teste Box'M.....	70
Tabela 22: Coeficientes de função discriminante canônica.....	70
Tabela 23: Teste de Wilks'Lambda.....	71
Tabela 24: Funções em centroides de grupo.....	71
Tabela 25: Probabilidades a priori para grupos.....	71
Tabela 26: Perfis dos segurados.....	72
Tabela 27: Resultados da classificação.....	73

LISTA DE QUADROS

Quadro 01: Sistema Nacional de Seguros..... 24

Quadro 02: Resumo sobre termos estatístico..... 31

LISTA DE ABREVIATURA E SIGLAS

SUSEP-	Superintendência de Seguros Privados
QAR-	Questionário de Avaliação de Risco
CNSP-	Conselho Nacional de Seguros Privados
FENASEG-	Federação Nacional das Seguradoras
AD-	Análise Discriminante
SPSS-	Statistical Package for the Social Sciences
MDA-	Multiple-group discriminantanalysis
FUNENSEG -	Fundação Escola Nacional de Seguros
PIB-	Produto Interno Bruto
VI-	Variáveis Independentes
VD-	Variáveis Dependentes

SUMÁRIO

1. INTRODUÇÃO.....	14
1.1 Problema da Pesquisa.....	16
1.2 Objetivos.....	17
1.3 Justificativa.....	17
1.4 Estrutura da Dissertação.....	18
2. REVISÃO BIBLIOGRÁFICA.....	19
2.1 REFERENCIAL TEÓRICO BÁSICO.....	19
2.1.1 A história do seguro no Brasil.....	19
2.1.1.1 Sistema nacional de seguros.....	23
2.1.1.2 Diferença entre risco e incerteza em seguros.....	24
2.1.1.3 A operação e a oferta do seguro.....	25
2.1.1.4 O papel de intermediador do corretor de seguros.....	27
2.1.1.5 Relevância do Seguro de automóvel.....	29
2.1.1.6 O Questionário de Avaliação de Risco (QAR).....	30
2.1.2 Análise Multivariada: Análise Discriminante.....	31
2.1.2.1 Resumo sobre termos estatísticos.....	31
2.1.2.2 Conceituação do modelo estatístico.....	32
2.1.2.3 Orientações e suposições para aplicação da análise discriminante.....	34
2.1.2.4 Caso de dois grupos.....	36
2.1.2.5 Caso de dois grupos com mais de uma variável independente.....	38
2.1.2.6 Caminhos para a busca da função discriminante.....	40
2.1.2.7 Testes estatísticos na análise discriminante.....	41
2.1.2.7.1 Método <i>Stepwise</i>	41
2.1.2.7.2 Testes para inclusão ou retirada de variáveis independentes.....	42
2.1.2.7.3 Teste para igualdade de matrizes de covariância.....	44
2.1.2.7.4 Testes para avaliar a qualidade do ajuste da função discriminante.....	45
2.1.2.7.5 Teste de chances.....	47
2.2 REVISÃO DE LITERATURA.....	47

2.2.1 Seguros de automóvel.....	47
2.2.2 Análise discriminante na identificação de perfis.....	48
3. MÉTODO DE PESQUISA.....	50
3.1 Pesquisa.....	50
3.1.1 Amostra.....	50
3.2 Etapas.....	50
3.2.1 Tratamento das potenciais questões.....	52
3.2.2 Análises estatísticas dos dados.....	52
3.2.3 Divisão da amostra total para construção do modelo.....	54
3.2.4 Método para construção do modelo.....	54
3.2.5 Verificação do modelo.....	54
4. APRESENTAÇÃO DOS RESULTADOS.....	55
4.1 Descrições das potenciais variáveis explicativas.....	55
4.2 Características dos segurados da amostra total.....	57
4.3 Seleção das variáveis.....	58
4.3.1 Testes para as variáveis.....	58
4.3.2 Amostra de desenvolvimento e de teste.....	68
4.3.3 Variáveis potenciais para a construção do modelo.....	69
4.3.4 Validação da função discriminante.....	73
5. CONCLUSÃO.....	75
REFERÊNCIAS BIBLIOGRÁFICAS.....	77
ANEXOS.....	84

1. INTRODUÇÃO

As pessoas estão sempre procurando caminhos que as conduzam a um modelo de vida onde os momentos de sucesso, felicidade e alegria superem, de modo significativo, aqueles de tristeza, sofrimento e dor. De qualquer forma problemas como desemprego, violência, saúde e outros estão na trajetória das pessoas e devem ser enfrentados. Para Gilberto (2010) são eventos aleatórios que chegam de surpresa e causam transtornos, não somente aos diretamente atingidos como também aos familiares, gerando, entre outros resultados, custos muitas vezes não programados.

Felizmente algumas potenciais ocorrências podem ser, no mínimo, amenizadas com prevenções planejadas. Dentre elas destacam-se os programas de seguros, tais como seguro de vida, saúde, residencial, empresarial e de automóvel - o qual oferece variados modos de cobertura, destacando danos por colisão, incêndio, roubo e contra prejuízos a terceiros. No caso dos veículos, os custos ficam na dependência de variáveis tais como, idade e modelo do veículo, itens de segurança, região mais frequente de uso do automotor, montantes das coberturas e características ligadas diretamente ao condutor, dentre elas idade, sexo e envolvimento em ocorrências que geraram acidentes (SILVA, 2007).

O mercado de seguros surgiu por volta do século XIX, em um período de grandes mudanças, que apontava grandes evidências de crescimento da economia, com o aumento dos postos de trabalho e de uma maior geração de renda; efeitos determinantes e mais positivos de um cenário progressista, seguindo em linha com a maior procura por seguros (PEREIRA, 2004).

Por volta de 1939 criou-se um efetivo Mercado Segurador Nacional, explica Bernstein (1997), o que fez crescer consideravelmente o número de seguradoras nacionais e obrigou as estrangeiras a se organizarem de acordo com as leis brasileiras.

No seguro de automóvel o contrato envolve, para a sua formação, a existência do segurado, o qual é representado por um corretor de seguros legalmente habilitado e a seguradora. Envolve ainda, o objeto do seguro, o prêmio que é pago pelo segurado e a indenização devida paga pela seguradora em caso de sinistros (acidentes). Segundo a Susep (2014) o segurado poderá ser toda pessoa física ou jurídica que contratar um seguro, transferindo para uma seguradora por meio de

uma corretora ou não, o risco de um possível evento atingir seu(s) bem(s) de interesse, para tanto se faz um pagamento ao segurador (valor pago pelo segurado, denominado de prêmio).

Esse prêmio é estimado em função da mutualidade dos riscos e do comportamento histórico do segurado em vista do seu envolvimento em ocorrências, as quais geram sinistros que representam a consumação dos riscos. Segundo Bernstein (1997) o risco está associado ao desconhecimento do futuro em vista dos possíveis acontecimentos, porém a sociedade vem estudando e aprendendo a lidar com ele, desenvolvendo, para tanto, métodos que, em conjunto com técnicas estatísticas, possibilitam procedimentos para gerenciá-lo. Portanto o seguro existe como forma de prevenir riscos indesejáveis, possíveis e incertos, em vista das circunstâncias da vida. É uma forma de ressarcimento dos efeitos de eventos casuais por meio de um fundo constituído pelo grupo de segurados a partir de leis da estatística, tal como a Lei dos Grandes Números enunciada por Bernoulli (MEYER, 1983).

Evidente que esses métodos devem levar às formulações equilibradas que garantam a “sobrevivência” da seguradora e não onerem de forma significativa o segurado. Importante neste sistema é o relato histórico dos riscos envolvidos. Isto é, quando fatos mudam a frequência desses riscos, custos adicionais podem ocorrer no valor do prêmio. Exemplificando: para o caso de veículos, cidades ou regiões onde a frequência de roubos é maior e/ou o sistema viário leva a chances maiores de acidentes, as seguradoras tendem a oferecer seguros com prêmios maiores (PAUZEIRO, 2007).

Para Ribeiro (2008) e Ramos (2010), em seus estudos é identificado que a maioria dos acidentes de trânsito envolve jovens de 18 a 24 anos. O que sugere que esta faixa etária seja entendida como mais propensa a ter sinistros.

Em relação especificamente ao condutor, seu perfil também é um componente importante no valor que despenderá para adquirir o seguro. Quando se fala em informações pessoais e outras, deve-se entender que, no ato da contratação, o cliente deverá escolher entre um contrato que tenha ou não cláusula de perfil - formulário onde se preenche as informações pessoais e outras do condutor principal (que não foi o solicitante do contrato, mas que representa quem mais utiliza o veículo) para cálculo do risco e seu respectivo prêmio.

Em resumo, Silva (2008) relata que o entendimento e a convivência com o risco foi o que originou a indústria de seguros.

O contrato de seguros de automóvel, assim como os demais contratos, por lei devem conter informações verídicas, para que tais dados pessoais e outros sejam passíveis de análise e deduções corretas. Desta forma, entende-se que em seguros, quando houver dados inválidos a seguradora poderá negar a indenização e atendimento ao segurado (OLIVEIRA, 2005).

Para se classificar o indivíduo alocando-o entre os sinistrados e não sinistrados pode-se utilizar as técnicas estatísticas como as descritivas e probabilísticas acompanhadas de outras mais sofisticadas como os Testes comparativos e a Análise Discriminante (AD). Mingoti (2005) define AD como uma técnica multivariada para a classificação de elementos de uma amostra ou população em categorias. No caso deste trabalho duas categorias (grupos) são definidas: dos condutores sinistrados e não sinistrados que compõem a variável dependente. A classificação é feita através da identificação das variáveis explicativas que possibilitam a segregação dos segurados pesquisados e dos futuros compradores de seguro automóvel, permitindo uma análise por parte da seguradora que possibilite gerenciar prêmios em função do perfil do cliente.

1.1 PROBLEMA DA PESQUISA

A classificação de um indivíduo, ou seja, a identificação da propensão para um sinistro (acidente) baseia-se nas informações recolhidas no ato da assinatura do contrato, momento em que o questionário de avaliação de risco (QAR - Anexo 1) é preenchido. Com base no perfil do segurado, como que uma seguradora pode identificar as variáveis que mais influenciam no sinistro e separar os indivíduos entre os grupos sinistrado ou não sinistrado?

Este trabalho tem a finalidade de construir um modelo estatístico que permita identificar as variáveis que possibilitam definir o risco associado aos segurados estudados e aos futuros compradores de seguros, evitando possivelmente o desequilíbrio financeiro das seguradoras.

1.2 OBJETIVOS

Objetivo Geral

Desenvolver um estudo por meio de um rol de informações sobre as características pessoais dos compradores de seguro de automóvel de uma corretora da cidade de Bauru, a fim de identificar as variáveis que possibilitam a segregação dos segurados pesquisados e dos futuros compradores de seguro de automóvel, permitindo a classificação desses potenciais/segurados nos grupos sinistrados e não sinistrados.

Objetivos Específicos

- Descrever as variáveis presentes nos questionários de avaliação de risco (QAR).
- Estratificar os dados em dois conjuntos, sendo um para a construção da função discriminante simples e outro para o teste desta função.
- Aplicar a Análise Discriminante determinando uma função discriminante
- Identificar quais as características (variáveis) que melhor discriminam os grupos.

1.3 JUSTIFICATIVA

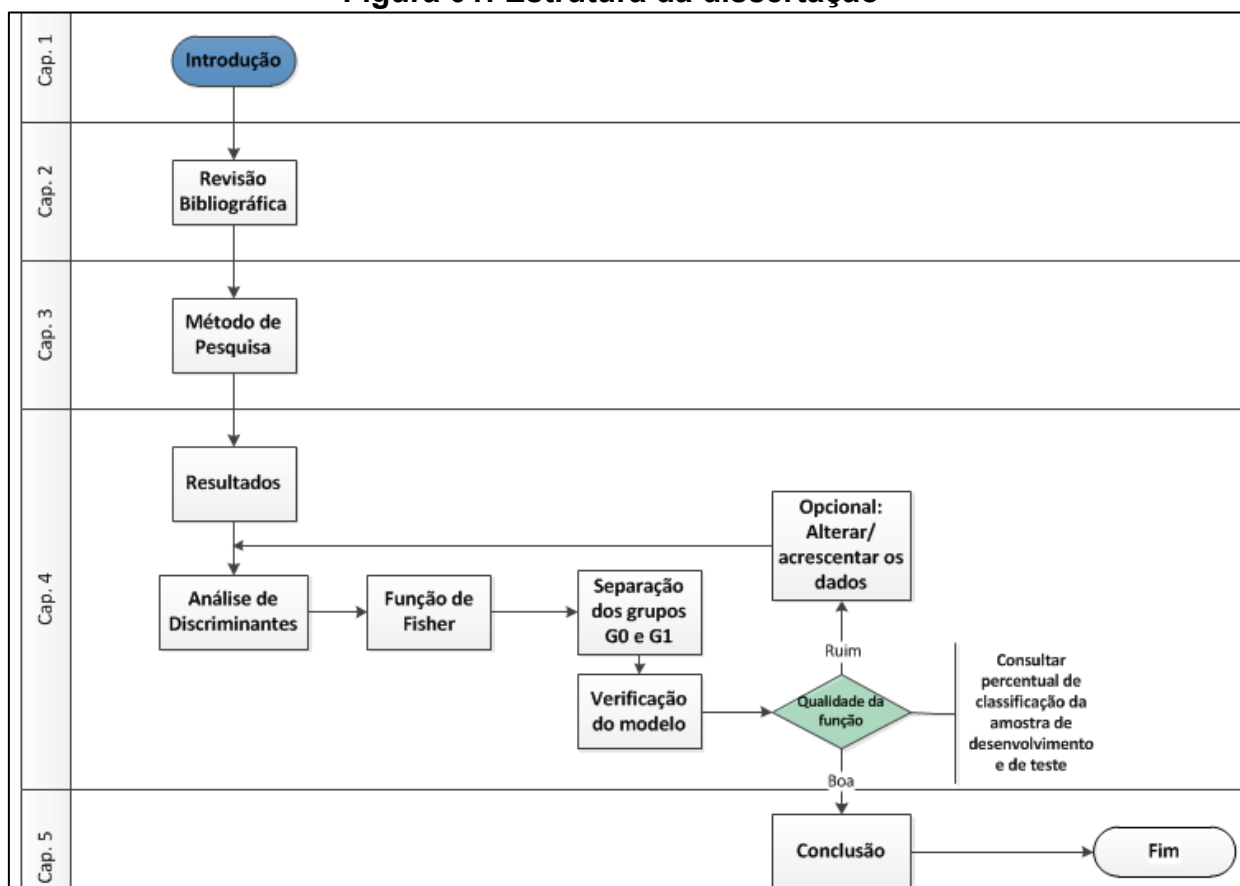
“Os seguros são uma pedra basilar na vida moderna. Sem seguros, segmentos da nossa sociedade e economia não funcionariam” (Insurance, 2012, p. 3). Partindo desta afirmação pode-se acrescentar que o grau de importância do seguro aumenta conforme a sociedade evolui em tecnologia e informação. Em geral, a indústria seguradora, desde o século XIX, tem proporcionado à sociedade e à economia um melhor grau de segurança, minimizando riscos e incertezas (Gilberto, 2010).

Portanto, para uma seguradora é importante uma correta avaliação do risco oferecido pelo potencial segurado, sob a pena de perda de recursos resultando em um mau funcionamento do mercado como um todo. Observa-se que, pela experiência prévia da autora, as seguradoras têm critérios distintos para análise do risco do condutor, porém todas utilizam as mesmas variáveis com ponderações diferenciadas. Neste contexto, a relevância desse estudo está na apresentação de uma proposta de modelo que identifique as variáveis e possibilite a determinação do risco associado aos segurados da amostra e dos futuros compradores de seguros.

1.4 ESTRUTURA DA DISSERTAÇÃO

Este trabalho está estruturado conforme sequência mostrada na Figura 01:

Figura 01: Estrutura da dissertação



No capítulo 1, a Introdução apresenta um texto breve sobre o tema da dissertação, identificação do problema, objetivos a alcançar e a justificativa de sua escolha.

O capítulo 2 trata da revisão bibliográfica, apresentando os conceitos de sinistralidade, seguro de automóvel, análise discriminante. O capítulo 3 descreve a metodologia de pesquisa. No 4º são apresentados os resultados obtidos e discussões. Por fim, o capítulo 5 apresenta as considerações finais.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta, através da revisão bibliográfica, os principais fundamentos necessários para o estudo dos elementos que serão abordados nesta pesquisa como: história, termos técnicos de seguros, riscos e incertezas e análise discriminante. Os artigos deste capítulo foram pesquisados a fim de se compreender e interpretar produções feitas anteriormente, com o uso da análise multivariada dentro e fora do ramo de seguros, visto o baixo número de trabalhos acadêmicos com este tema, a fim de melhorar e propor novas estratégias de segregação de perfis de segurados.

2.1 REFERENCIAL TEÓRICO BÁSICO

Esta sessão traz o referencial teórico básico sobre tópicos relativos ao seguro de automóvel; perfil do segurado, sinistros e previsão e Análise Discriminante.

2.1.1 A história do seguro no Brasil

Martins (2005) cita que o termo seguro vem do latim e que significa livre de angústia, sem inquietação. O nascimento do seguro acompanha a história da humanidade, que desde o princípio da civilização, de certa forma, já se protegia contra o perigo, do medo dos acontecimentos imprevisíveis e da perda de seus bens materiais.

Segundo Ribeiro (1994) desde o início da civilização os povos já se protegiam contra eventos futuros, como naufrágios. Para tanto, evitavam incluir uma mercadoria em uma única embarcação. E, ao passo que os trabalhadores de grandes obras trabalhavam em conjunto, constituíam a chamada Caixa de Auxílio Mutua para socorrer e auxiliar seus familiares em caso de acidentes. Funcionava como uma arrecadação em dinheiro para o usufruto de seus familiares em casos de sinistros.

O mercado de seguros surgiu por volta do século XIX, em um período de grandes mudanças, que apontava grandes evidências de crescimento da economia, com o aumento dos postos de trabalho e de uma maior geração de renda, efeitos

determinantes e mais positivos de um cenário progressista, seguindo em linha com a maior procura por seguros (PEREIRA, 2004).

Para Bernstein (1997) a atividade de seguros chegou junto com a abertura dos portos ao comércio internacional, isso por volta de 1808. Nesta mesma época, surgiu a primeira sociedade de seguros chamada de "Companhia de Seguros BOA-FÉ", tendo por objetivo operar no setor marítimo.

Em meados de 1850, segundo Galiza (2000) a atividade seguradora deixou de ser coordenada pelas leis portuguesas passando a ser, pela primeira vez, estudada e regulada no Brasil. O "Código Comercial Brasileiro" teve grande influência no desenvolvimento do seguro no Brasil, estimulando a criação de inúmeras seguradoras que começaram a trabalhar com seguro terrestre além do marítimo. E no ano de 1855, o seguro de vida, antes proibido, foi autorizado e passou também a ser explorado.

Segundo Galiza, (2000) a regulamentação dos seguros no Brasil iniciou-se em 1860 e foi em 1901 que surgiu a primeira lei, pelo Decreto-Lei nº 4.270, Regulamento Murinho, que ordenava as transições de seguros e formalizava a criação de uma Superintendência Geral de Seguros.

Em 1º de janeiro de 1916, foi publicado o Código Civil Brasileiro, com uma seção que discriminava e direcionava os seguros em geral, exceto os Marítimos, já então regulados pelo Código Comercial (OLIVEIRA, 2005).

Por volta de 1939 criou-se um efetivo Mercado Segurador Nacional, explica Bernstein (1997), o que fez crescer consideravelmente o número de seguradoras nacionais e obrigou as estrangeiras a se organizarem de acordo com as leis brasileiras.

O crescimento deste setor fez com que o interesse das empresas estrangeiras aumentasse em relação ao mercado brasileiro, tanto que surgia no Brasil as primeiras sucursais (filiais) das seguradoras do exterior. E, em 5 de setembro de 1895, foi promulgada a Lei nº 294, que protegendo o país impedia que estas sucursais enviassem os recursos financeiros obtidos para suas matrizes. Esta lei garantiu que os recursos fossem aplicados no Brasil para fazer frente aos riscos aqui assumidos. Há registros de que algumas empresas não aceitaram esta regra e fecharam suas sucursais (ALVIM, 1999).

“A palavra risco está ligada aos termos latinos *risicu* e *riscu*, ligados por sua vez a *resicare*, que significa corte. Como uma ruptura na continuidade, como um

risco que se faz numa tela em branco” (MONTEIRO 1991, p. 10) citada por (LIMA 2010, p. 13).

Por volta de 1966, o Sistema Nacional de Seguros Privados começou a regular as operações de seguros e resseguros (seguro do seguro já contratado). E, neste mesmo período, surge o Conselho Nacional de Seguros Privados (CNSP) e a Superintendência de Seguros Privados (SUSEP), destaca Magalhães (1997). O poder regulador do Estado ampliou-se e passou a contar com uma grande procura dos bancos como consumidores de seguros.

E, desde então, a atividade seguradora tem sido essencial para pessoas físicas e jurídicas que decidem se precaver de riscos futuros. Segundo a Federação Nacional das Seguradoras – FENASEG, em 2013 a contribuição do setor de seguros no PIB foi próximo a 5,7% e, ainda declara que, historicamente, o mercado de seguros brasileiro tem crescido cerca de 3%, em média, acima do PIB. Segundo as Estatísticas em 25 anos, o mercado dobrará a participação em relação ao PIB.

No tocante ao ramo de automóveis, na Tabela 01 há termos usuais que devem ser entendidos como:

Tabela 01: Termos de Seguros

Termos	Definições
Seguro	Estrutura que reúne os riscos de um grupo de pessoas e suas contribuições para em caso de sinistro(s) existir provisões monetárias para pagar suas perdas vinculadas a esses riscos (PEREIRA, 2004).
Seguradora	Pessoa jurídica que assume a responsabilidade dos riscos do segurado (ALVIM, 1999).
Segurado	Pessoa física ou jurídica em nome da qual o seguro é contratado (SANTOS, 1944). TM
Corretor de Seguros	É o intermediário, pessoa física ou jurídica, entre segurado e a seguradora (BULCÃO, 1989).
Questionário de avaliação	Formulário que contém perguntas à respeito do bem

de risco (QAR)- Anexo II	segurado (veículo), do condutor e das coberturas contratadas (SILVA, 2007).
Perfil	Conjunto de perguntas que a seguradora faz para definir o perfil do segurado. E, com base nestas respostas, concedem descontos ou agravam o valor do seguro contratado (ALVIM, 1999).
Risco	Evento possível, futuro e incerto, independente da vontade do segurado, cuja ocorrência gere prejuízos financeiros (GALIZA, 2000).
Sinistro	Evento não planejado e imprevisível, acidente que causará danos ao bem(s) segurado(s) e que se coberto pelo seguro, deverá ser indenizado pela seguradora (BULCÃO, 1989).
Custo do risco	Valor do sinistro, análise matemática através de uma avaliação de três variáveis: frequência do risco; severidade (custo monetário); dor e sofrimento humano (PAUZEIRO, 2007).
Importância segurada	Valor que limita a responsabilidade da seguradora (SILVA, 2007).
Franquia	Valor máximo em dinheiro que o segurador não se responsabiliza a indenizar o segurado em caso de sinistro (PEREIRA, 2004).
Cobertura do risco	Representa os danos (tipos de sinistros) que serão indenizados (PINHO, 2003).
Prêmio	Valor pago no ato da contratação do seguro pelo segurado ao segurador (OLIVEIRA, 2005). Silva (2008) acrescenta que o valor do prêmio é baseado nas informações prestadas pelo mesmo no ato da assinatura da proposta, ao preencher o Questionário de Avaliação de Risco (QAR).
Mutualismo	O mutualismo está relacionado à união de esforços de muitos em favor de alguns elementos do grupo,

	já que estes, isoladamente, não teriam condições de suportar prejuízos de monta. As seguradoras administram os prêmios pagos pelos segurados e este montante servirá de indenização em caso de sinistro de um dos mesmos (SILVEIRA, 2008).
--	--

Fonte: Elaborado pelo autor

Para Alvim (1999) citado por Costa (2013) a contratação do seguro de automóvel com base no QAR assegurará cobertura de certo bem em relação à determinada(s) pessoa(s).

A probabilidade de ocorrência de sinistro pode perder seu caráter aleatório quando o segurado fornece informações falsas ou indevidas no QAR, ou seja, quando a qualidade de informação é pouca entre o indivíduo e a seguradora menciona Dionne et. al. (1999).

Para Chiappori et. al. (2000) a qualidade da informação é um fator muito importante para a indústria seguradora, pois é ela que permite avaliar e julgar o grau de risco de quem pretende comprar um seguro. E este gargalo permite que “maus condutores” tenham vantagens no seguro de seu risco.

Partindo desse pressuposto, a companhia de seguros antes de celebrar um contrato precisa ter métodos ou um conjunto de ferramentas que lhe permita avaliar os diferentes níveis de risco entre seus vários clientes, visto que a partir da contratação do seguro, o segurado pode vir a assumir atitudes e comportamentos mais arriscados que antes não os tinha. Isto porque, inconscientemente, transfere a sua responsabilidade para outro.

2.1.1.1 Sistema nacional de seguros

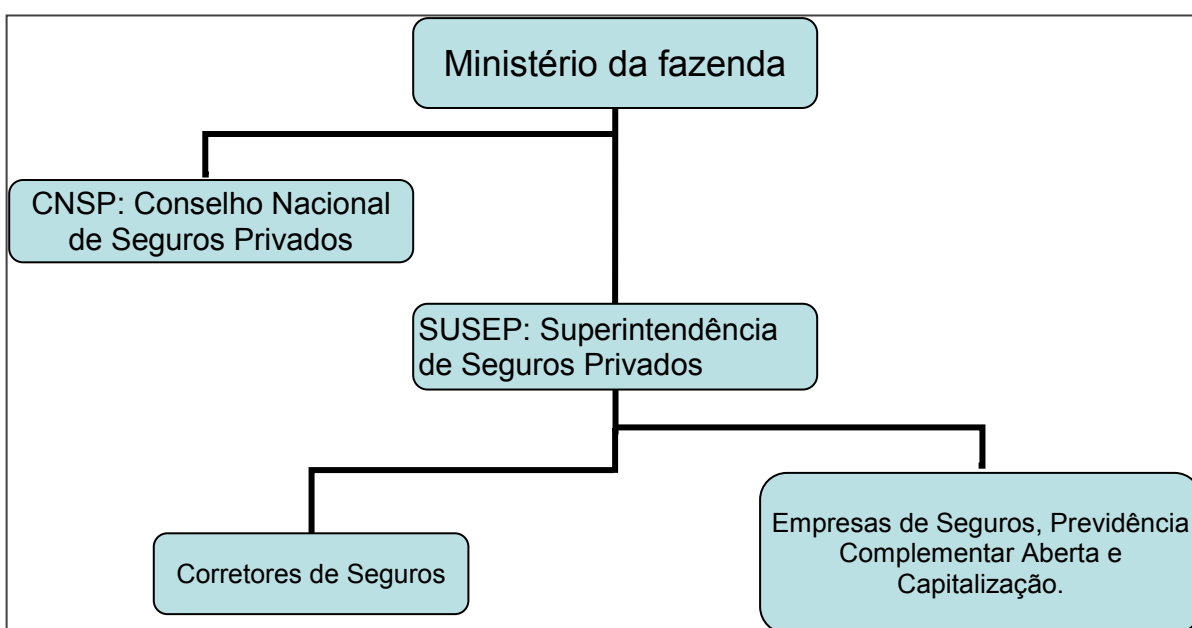
A Federação Nacional das Empresas de Seguros Privados e de Capitalização –FENASEG - é uma associação sindical fundada em 1951 com sede e foro na cidade do Rio de Janeiro, com a obrigação de estudar, coordenar, proteger e representar as categorias do seguro privado (OLIVEIRA, 2005).

De acordo com Magalhães (1997), o papel da SUSEP é o de habilitar corretores que se submetem a uma prova de capacidade técnica profissional

realizada pela Fundação Nacional de Seguro – FUNENSEG, entidade criada para promover a qualificação dos indivíduos que trabalham com seguros.

Ainda Magalhães, menciona que a Federação Nacional de Corretores de Seguros e de capitalização (FENACOR) é uma entidade que representa judicialmente os sindicatos afiliados a ela. Sindicados estes que representam os corretores de seguros. Com base no informe Fenaseg (2007), o Sistema Nacional de Seguros Privados (SNSP) do Brasil pode ser compreendido como no Quadro 01:

Quadro 01: Sistema Nacional de Seguros



Fonte: Adaptado de Fenaseg 2002

Para Fenaseg (2007), a tarefa de formular políticas de seguros privados e de fiscalizar as operações no mercado nacional estão sob a responsabilidade do Governo Federal.

2.1.1.2 Diferença entre risco e incerteza em seguros

A sociedade vive em constante evolução em meio a diversos riscos. Para Larramendi (1982), apesar dos aspectos coloquiais tratarem risco e incerteza como tendo um mesmo significado, em termos técnicos isto se destoa. O termo risco tem em sua particularidade o significado de um grau de incerteza de perda de valor econômico. Ambos os termos, resultam em grande parte de não se poder prever o

futuro com segurança; o risco se trata de um evento aleatório que acarretará em possíveis perdas, de ordem material, pessoal ou mesmo de responsabilidade. Em contrapartida, a incerteza é algo não evidente, ou seja, que não pode ser estimado.

Ainda na visão de Sanvicente (1997, p. 64), “a incerteza e o risco decorrem, em grande parte, da impossibilidade de prever o futuro com absoluta segurança”.

E, para Figueiredo (1997, p. 19), “risco é uma condição em que perdas são possíveis”.

Em geral, quando se conhece a variável estudada (ocorrências, variações futuras...) e desta pode-se estimar uma probabilidade, diz-se que existe risco. Agora, quando não é possível se obter uma probabilidade, diz-se que há incerteza. A incerteza envolve situações não repetitivas ou pouco comuns, cujas probabilidades não podem ser determinadas (LARRAMENDI, 1982).

De forma simples, Oliveira (2005) detalha que o risco pode ser medido e analisado com base nas distribuições probabilísticas estimadas, em torno do histórico de dados. A seguradora poderá, desta forma, estimar probabilidades relativas aos vários resultados obtidos. Na ausência de dados históricos, com estimativas aceitáveis e subjetivas, existe a incerteza.

2.1.1.3 A operação e a oferta do seguro

O texto de Mauss (1974) apresenta uma visão de que a mutualidade é necessária para que, tanto indivíduos quanto a sociedade, não obtenham prejuízos irreparáveis, para tanto com base em probabilidades a seguradora pode estimar futuros riscos.

Para Bulcão (1989) risco é um evento que causará danos a indivíduos e a bens, por exemplo: uma colisão de veículos. Neste sinistro, o culpado terá a obrigação de ressarcir o prejuízo causado à parte envolvida (terceiro no sinistro). Desta maneira, a seguradora irá colaborar para um equilíbrio econômico entre os envolvidos, pagando os danos causados pelo seu segurado.

Para Viola (1983) o seguro sempre envolverá a incerteza de ocorrências de sinistros, também sempre terá seu caráter de proteção principalmente quando o termo Previdência for utilizado. E o mutualismo será a ferramenta que poderá

influenciar na economia social do país, partindo da interpretação que os prêmios (pagamentos) de um grupo de pessoas serão utilizados para benefício dos mesmos.

No contexto histórico, segundo Pinho (2003), também pode-se dizer que o mutualismo reduz o risco e advém de definição bíblica contida em “O Novo Testamento” (Gálatas, 6:2), que diz: “Levai as cargas uns dos outros”.

A boa fé pode ser compreendida como um princípio de boa índole e honestidade, visto que cada envolvido: segurado, corretor e seguradora devem usar de informações reais e verídicas em todo o processo de contratação do seguro. E, por saber que as informações omitidas ou errôneas podem ocasionar transtornos e desequilíbrio para ambas as partes: corretor, segurado e segurador (LUGON et. al. 2004).

Em um de seus artigos Lugon et. al. 2004, trabalha também a questão da análise do risco com o apoio da análise discriminante, sua amostra inicial partiu de 32.000 (trinta e dois mil registros), sendo que deste total apenas 21.024 foram considerados completos para a análise (30% não sinistrados e 70% sinistrados). O modelo trabalhou com 15 variáveis, entre algumas *dummys*. Terminando a análise com 9 variáveis, consideradas as mais discriminantes, mas que classificaram muito pouco os casos, aproximadamente 54% dos segurados.

Martins (2004) define que seguradora é a pessoa jurídica que gerencia e responde pela aceitação do risco. A necessidade de um cliente transferir seu risco a mesma, perante a legislação brasileira, depende do intermédio de um corretor, que quando escolhido poderá apresentar uma cotação (orçamento) contendo o preço do seguro ora ofertado. Caso o cliente conheça um pouco mais da operação de seguros, também poderá solicitar explicações sobre coberturas e limites de responsabilidade, ofertadas pela(s) seguradora(s) ora consultada(s) e/ou recomendada(s) pelo corretor de seguros. Inicia-se, assim, o processo de cotação/elaboração da proposta de seguros para o cliente.

A elaboração de uma cotação/proposta de seguro é uma tarefa árdua, que envolve em seu processo inúmeras questões e um grande investimento em tecnologia, processos, gestão e marketing, tanto das seguradoras quanto dos corretores de seguros. É necessário ainda que a seguradora possua acesso a um banco de dados (próprio ou da corretora) contendo as informações do segurado e do risco a ser coberto para que possa precificar o seguro ao corretor, representante do cliente. Desta forma, quanto maior a quantidade de dados a serem analisados, maior

será a complexidade da análise e mais criteriosa será a estratégia no momento de ofertar o seguro (CASTIGLIONE, 1997).

Para Silva (1984) durante a precificação a seguradora estará verificando quais serão os riscos que deverá aceitar e/ou quais riscos deverão ser recusados. A decisão da seguradora ainda estará na redução ou acréscimo do prêmio. Riscos com menor sinistralidade tendem a ter um menor prêmio de seguro. Do mesmo modo, os riscos com maior sinistralidade tendem a ter um maior prêmio. Nesta fase, o histórico do segurado muito influenciará na aceitação e precificação do seguro.

Lemaire (1995) confirma a tese de Silva (1984) descrevendo que para agravar ou não o valor do prêmio de um seguro, de um ano para o outro as seguradoras utilizam um sistema chamado de bônus-malus, que resume-se a diminuir (bônus) o valor pago caso não se registre sinistros ou a agravá-lo (malus) caso se registre um ou mais sinistros.

Em geral, é cada vez mais difícil para as seguradoras manterem diferenças de preços entre as muitas categorias de risco deste mercado competitivo. Cientes disto, a criação de uma ferramenta tarifária que equalize o custo dos sinistros pelos segurados estimula a procura de fatores de classificações adicionais, ou seja, características e informações, até o momento não exigidas, para precificação ou simples análise, mas que poderão ser determinantes em um minuto de competição entre seguradoras (DENUIT et. al. 2007).

Assim, com a grande concorrência existente no mercado segurador há clientes que facilmente mudam de companhia de seguros à procura de maiores descontos e melhores condições que se traduzam num preço mais reduzido para o seu seguro.

2.1.1.4 O papel de intermediador do corretor de seguros

De acordo com Martins (2004), a intermediação em qualquer tipo de negócio normalmente é vista como fator de encarecimento do produto ou do serviço a ser prestado. Todavia, não se tem o exato entendimento do papel de quem faz a intermediação. Tratando-se de seguro, pela própria natureza do negócio, a intermediação é primordial. Em primeiro lugar, o contrato é complexo, gerando a necessidade de o comprador de seguro ter esclarecimentos sobre a sua natureza e adequação desta natureza às suas necessidades, tarefa essa que exige

conhecimento das particularidades do mercado e do negócio e, conseqüentemente, de certa perícia.

Segundo Castiglione (1997), antes de surgir a figura de um corretor, as propostas (documento que apresenta um orçamento do seguro) eram fornecidas como meros formulários padrões. Era comum as apólices (documento legal que firma o compromisso entre as partes, com base na proposta aceita pelo segurado) possuírem sempre as mesmas particularidades. Desta forma, surgiu a função do corretor que, desenvolvendo técnicas de vendas e se aprimorando em conceitos, passa a ser um assessor de seus clientes, fornecendo benefícios contratuais e, acima de tudo, segurança aos compradores de seguros.

Assim, Silva (1984) cita que na intermediação do contrato de seguro o cliente é direcionado pelas informações fornecidas pelo corretor que está intermediando o negócio e, através destas, o segurado aceita ou não o contrato de seguro. O corretor se torna o representante legal do segurado para todas e quaisquer tratativas que se relacionem com o bom fechamento do contrato. Na existência de uma falha durante a prestação do serviço por parte do corretor, este responderá pelos prejuízos que vier a causar, seja perante a seguradora ou ao segurado.

Bulcão (1989) afirma que o seguro não é um produto que um indivíduo adquira tendo a mesma ideia de que está adquirindo uma mercadoria qualquer. O seguro exige esforço para uma compreensão de custos, deveres e obrigações, vantagens e desvantagens que terá o futuro segurado. Normalmente, ninguém se anima ao adquirir um produto que está ligado a um evento que lhe traga tristeza e/ou uma perda financeira, daí a necessidade de intermediação.

E Martins (2004) acrescenta que a lei brasileira permite ao cliente a escolha de um intermediário para a contratação de seu seguro. No mercado brasileiro é incomum o cliente não escolher por um corretor e contratar diretamente com uma seguradora.

De acordo com a Legislação Federal do Brasil o art. 18 da Lei nº 4594/64 regulariza a profissão do corretor de seguros, permitindo às seguradoras o recebimento de proposta de contrato de seguros por intermédio do mesmo. No Brasil, a compra de seguros via corretor não é obrigatória, mas aconselhável visto a complexidade dos contratos e negociações.

Segundo o Centro de qualificação do corretor de seguros - CQCS (2014), o superintendente da Susep Roberto Westenberger declarou que o corretor de seguros deve ser reconhecido e que, para tanto, deve-se valorizar tal autarquia. Assim, apontou que em sua gestão estes terão reconhecimento e voz para decisões que beneficiem todo o setor.

2.1.1.5 Relevância do seguro de automóvel

“O seguro de automóveis tem como função cobrir perdas ou danos dos veículos terrestres de propulsão a motor e a seus reboques, desde que não trafeguem sobre trilhos” (SOUZA, 2007, p. 65).

De acordo com o CQCS (2014), os sinistros de automóveis, com base em cálculos do Sindicato de Corretores do Estado de São Paulo, aumentaram em cerca de 30% a 40% no ano de 2014, comparado ao mesmo período de 2013. Em Sorocaba, por exemplo, entre roubos e furtos, no período do 1º quadrimestre de 2013, o total de casos atingiu uma soma de 1.198, enquanto que em 2014 atingiu um total de 1.303 sinistros.

Gameiro (2002) em seu artigo destaca que o roubo representa um forte empecilho para setores específicos como o de Transportes e para a sociedade como um todo.

Para Alvim (1999) o seguro de automóvel é parte dos oito mercados mais relevantes do ramo “não vida”, com um número considerável de companhias de seguros. As estatísticas mostram uma concentração do volume de negócios dentro de um número relativamente reduzido de seguradoras.

O aumento de clientes em uma corretora pode ser alcançado com a venda dos seguros de automóvel. Esta é uma estratégia muito utilizada e como o preço é um dos fatores mais importantes para o cliente, apresentar o preço mais atrativo (mais baixo) faz a diferença. Outra sugestão será ousar em coberturas diferenciadas e em um bom relacionamento para cativar o cliente (WILLIAMS, 1985).

2.1.1.6 O Questionário de Avaliação de Risco (QAR)

Segundo a Susep, o Questionário de Avaliação de Risco – QAR pode ser compreendido como um formulário que contém uma série de perguntas que a seguradora faz para analisar o grau de risco que ela irá assumir, caso o segurado concorde em contratar o seguro.

Há seguradoras que concedem descontos ou agravam (aumentam) o valor do prêmio de seguro de acordo com as respostas fornecidas pelo segurado. Dentre as perguntas, existem informações sobre o cliente (CPF ou CNPJ) que irá contratar o seguro, dados do objeto ou pessoa a ser consertado/ressarcido em caso de sinistro e detalhes de quais serão as coberturas contratadas (valor de ressarcimento de acordo com cada tipo de sinistro, exemplo: explosão, danos morais, morte, invalidez e etc).

No seguro de automóvel o QAR contém dados cadastrais: do indivíduo que contratará o seguro, dados do veículo e das informações do perfil (perguntas direcionadas à pessoa que utilizará por mais tempo o veículo segurado) como:

- Idade do principal condutor do veículo.
- Sexo do principal condutor do veículo.
- Região de Circulação do Veículo
- Uso de garagem ou estacionamento com frequência.
- Cargo
- Histórico de sinistros.

De acordo com o art. 766 do código civil, se o segurado por si só ou por seu representante fornecer declarações inválidas ou omitir circunstâncias que podem influenciar na aceitação da seguradora, perderá o direito a garantia de ressarcimento em caso de sinistro. Se identificado que a inexactidão ou omissão nas declarações não foram por má-fé do segurado, o segurador terá direito a corrigir o contrato e/ou a cobrar, mesmo após o sinistro, a diferença do prêmio.

A má fé, que no ramo de seguros corresponde a mentir intencionalmente no questionário para se pagar prêmio menor, pode acarretar na negativa de uma indenização.

Através da análise discriminante, com o uso das informações do QAR é possível identificar perfis de segurados que podem ser alocados no grupo de

sinistrados e não sinistrados. No tópico 2.2, são apresentadas as técnicas estatísticas que podem ser utilizadas em análises do mercado securitário.

2.1.2 Análise multivariada: Análise discriminante

2.1.2.1 Resumo sobre termos estatísticos

O Quadro 02 apresenta um resumo explanatório a respeito de termos estatísticos encontrados neste trabalho.

QUADRO 02: Resumo sobre termos estatísticos

Testes Estatísticos	Definições
População Normal	Para Hair et. al. (2005), a característica de uma distribuição normal é a sua forma de “curva do sino” quando vista graficamente. Sua função densidade tem apenas dois parâmetros: média e desvio padrão.
Nível de Significância	Probabilidade de se cometer erro do Tipo I, onde se rejeita a hipótese nula, quando a mesma é verdadeira (Mingoti, 2005).
P-value	Corresponde ao menor nível de significância que pode ser assumido para rejeitar a hipótese nula. Há significância quando o valor de p-value é menor do que o nível de significância adotado (LATTIN, et. al. 2011).
U - Wilks'Lambda	De acordo com Hair et. al. (2005), este teste multivariado avalia as diferenças de médias para os casos de múltiplas variáveis dependentes intervalares e múltiplos grupos (mais de dois) formados pelas variáveis independentes. O teste t e Z, e o teste F também tem esta função na análise univariada.
“T” de Student	Segundo Mingoti (2005), este teste conhecido como teste de Student, ou simplesmente teste t é utilizado frequentemente para se avaliar diferenças entre as médias de dois grupos. Ele é utilizado nas situações de pequenas amostras com desvio padrão populacional desconhecido.
Qui-Quadrado	Para Lattin, et. al. (2011), o teste Qui-quadrado é utilizado para verificar a significância da associação entre duas variáveis qualitativas A e C. Na Análise multivariada é utilizado para testar a significância da função discriminante.
M de Box	Teste para avaliar a homogeneidade das matrizes de covariância de variâncias e covariâncias. Desenvolvido por Box (1950), sendo uma generalização do teste univariado de igualdade de variâncias de Bartlett (1947).

Correlação	A correlação representa uma medida padronizada da relação entre duas variáveis. Nas medidas mais usuais ela não pode ser maior que 1 ou menor que -1. Se próxima a zero indica que as duas variáveis não estão relacionadas e, se próximas de 1 ou -1 indicam que elas estão fortemente correlacionadas. (MINGOTI, 2005).
Coeficiente de correlação linear de Pearson	Para Hair et. al. (2005), esta medida fornece um valor numérico que correspondente ao grau de correlação entre duas variáveis. No caso de duas variáveis contínuas é usual o coeficiente de correlação linear de Pearson o qual varia de -1 a 1.
Variância	É uma medida de dispersão que indica a distância média quadrática dos valores de uma série em relação à média aritmética desta série. A raiz quadrada da variância é o desvio padrão, mais usual nas comunicações pelo fato de figurar com a mesma unidade dos valores (JOHNSON e WICHERN, 1992).
Multicolinearidade	Quando uma variável pode ser explicada por outra entende-se que existe um tipo de interdependência. A essa relação de interdependência chamamos multicolinearidade (Fonseca, 2003). Duas ou mais variáveis altamente correlacionadas.
Centróides	Segundo Mingoti (2005), os centróides representam o valor médio para os scores K discriminantes de todos os objetos, em um grupo.
Q Press	Para Hair et. al. (2005), o Q Press é teste que valida o poder de discriminação da matriz de classificação baseando-se no número das classificações corretas do modelo.

Fonte: Elaborado pelo autor

2.1.2.2 Conceituação do modelo estatístico

A Análise Discriminante (AD) é uma técnica multivariada utilizada para classificação de elementos de uma amostra ou população (MINGOTI, 2005), quando a variável dependente é categórica, isto é, qualitativa e as variáveis independentes são quantitativas (FÁVERO, et. al. 2009). Segundo esses autores, quando a AD envolve uma variável dependente com dois grupos é denominada Análise Discriminante Simples, quando o número de grupos é superior a dois, a denominação é Análise Discriminante Múltipla (MDA-Multiple-group Discriminant Analysis).

A AD é aplicável quando um conjunto de elementos (amostra ou população) pode ser separado (discriminado) em dois ou mais grupos considerados como possíveis valores de uma variável dependente não métrica (HAIR, et. al. 2005). Seu objetivo é, em função das informações das variáveis independentes, separar ou discriminar os indivíduos nos grupos não métricos do modo mais claro possível (LATTIN et. al, 2011).

Em resumo, a AD busca a separação (discriminação) de um conjunto de objetos (através de observações) e\ou alocação (classificação) de novos em grupos já estabelecidos (JOHNSON e WICHERN, 1992), ao passo que se analisa as variáveis (aspectos e características) nos objetos que fazem parte de cada grupo.

Ainda segundo esses autores, trata-se de um instrumento multiuso que possibilita a obtenção de muitas análises, tais como redução de dados ou na simplificação estrutural, classificação ou agrupamento, investigação de dependência entre variáveis, predição, elaboração e teste de hipóteses.

Esta técnica é capaz de gerar previsões a respeito de qual grupo pertencem pessoas ou itens, uma vez que se caracteriza como uma técnica de previsão de classificação. Para tanto, ela gera funções discriminantes que ampliam a segregação dos grupos descritos pelas categorias de determinada variável dependente (ESCOFIER e PAGES, 1992).

Uma instituição financeira pode ter interesse em classificar futuros clientes em duas categorias: bom e mau pagador (possível inadimplente) em função dos perfis de seus atuais clientes. Diversas informações (variáveis) dos indivíduos desses dois grupos são consideradas a fim de se caracterizar o bom e mau pagador. Em termos matemáticos, são construídas combinações lineares dessas variáveis dentro de cada grupo, gerando uma função discriminante que possibilita classificar indivíduos com maior probabilidade de acerto em um desses grupos. Em resumo, um novo tomador de empréstimo, ao apresentar seus dados (variáveis), tem os mesmos inseridos (valor numérico) em uma função matemática (discriminante) que define a qual grupo ele pertence e qual sua respectiva probabilidade. Isto possibilita à empresa definir procedimentos em relação ao indivíduo. De qualquer modo, deve-se estar atento para o fato de que o erro de decisão pode ocorrer (MINGOTI, 2005).

Segundo a autora, o objetivo da AD é construir uma regra que minimize a quantidade de classificações incorretas e o custo associado a elas. Evidente que, nem todas as variáveis anotadas sobre os indivíduos têm potencial para interferirem

na discriminação. Hair et. al (2005) afirma, em relação aos objetivos da AD, que uma exploração inicial deve ser realizada a fim de verificar quais das variáveis independentes têm de fato possibilidade de influir de forma significativa na discriminação e quais os respectivos pesos de suas participações. A AD permite destacar as variáveis mais influentes na discriminação utilizando testes estatísticos como o Lambda de Wilks e o Qui-quadrado.

Fávero et. al. (2009), Hair et. al (2005) e Lattin, et. al. (2011) relatam ainda que a AD tem como objetivo descrever os comportamentos das variáveis independentes (VI) em função dos grupos, testando estatisticamente em termos de médias as significâncias das diferenças entre as VI dentro dos grupos. Seu principal objetivo é estabelecer procedimentos para classificar objetos em grupos com base nas variáveis independentes.

2.1.2.3 Orientações e suposições para aplicação da análise discriminante

Na aplicação de muitas técnicas estatísticas, condições iniciais devem ser satisfeitas para que as aplicações apresentem resultados confiáveis. Para a realização da análise discriminante é necessário que exista uma normalidade multivariada (variáveis independentes escolhidas aleatoriamente de uma população com distribuição normal) e homogeneidade das matrizes de variância e covariância, ou seja, não deve existir multicolinearidade (duas ou mais variáveis independentes altamente similares) entre as variáveis, se houver não se recomenda o uso da AD (HAIR, 2005).

Na definição dos grupos categóricos (possíveis indicações da variável dependente), que podem ser dois ou mais, esses devem ser mutuamente excludentes e abranger todos os casos (HAIR et. al. 2005, FÁVERO, et. al. 2009, LATTIN, et. al. 2011). É possível que uma variável métrica seja considerada como dependente desde que convenientemente categorizada. Na definição das variáveis independentes (VI) podem ser consideradas informações anteriores ou até a intuição, a partir do conhecimento de especialistas. Ainda em relação as VI, como em outras análises estatísticas, a suposição da normalidade multivariada deve ser considerada juntamente com a questão da homogeneidade das matrizes de variância e covariância. Outra condição a se verificar é quanto a multicolinearidade.

Uma análise inicial das correlações pode auxiliar na eliminação de variáveis altamente correlacionadas. Em relação as VI, destaque-se ainda que na AD informações podem ser disponibilizadas em função de variáveis nominais, tais como sexo, estado civil. Isto contraria a condição inicial de que as VI devem ser métricas, porém Hair et. al. (2005) comenta sobre variáveis dicotômicas (0 e 1) atuando como variáveis de substituição ou codificação indicadora, possibilitando desta forma a utilização de variáveis nominais no estudo da AD. Para corrigir e auxiliar as pesquisas que envolvem dados qualitativos (sexo, estado civil e outros), tem-se a opção de transformá-las em variáveis *dummys*, contendo n-1 subvariáveis onde n é o número de níveis de cada variável.

Entende-se que uma variável *dummy* se trata de uma variável que representa uma situação qualitativa, também chamada de não métrica, como gênero, por exemplo, representada por duas variáveis:

$X_1 = 1$, se o gênero for feminino;

$X_2 = 0$, se o gênero for masculino.

Para Corrar et. al. (2012), *dummy* são conhecidas também como variáveis binárias ou categóricas utilizadas para apontar a presença ou ausência de algum atributo ou “qualidade”, como homem e mulher, assumindo apenas os valores de 0 ou 1. A forma de se mensurar estes atributos é criar variáveis artificiais numéricas que apontam para elas.

Em relação ao tamanho da amostra total, os autores (HAIR, et. al. 2005) relatam sobre informações empíricas que ajudam a definir essa dimensão. Os autores recomendam que cada grupo possua, no mínimo, 20 observações e os tamanhos relativos também às observações devem ser considerados, pois grupos maiores têm chances maiores de classificação.

Para Ferreira (2008) e Hair, et. al. (2005) um conjunto de dados pode ser dividido em duas partes, uma para se obter a construção da(s) função (ões) discriminante(s) (parte da amostra chamada de desenvolvimento) e outra para testar a função (parte da amostra chamada de teste). Hair, et. al. (2005) afirma que o mais usual é repartir igualmente a amostra, mas alguns pesquisadores preferem partições com subamostras maiores para os grupos de análise, tais como 60% e 40% ou mais

extremo, 75% e 25%. Evidente que, para essa divisão, a sugestão é que se tenha no mínimo 100 elementos para análise.

2.1.2.4 Caso de dois grupos

Tanto para os casos mais simples, com número de grupos (g) igual a dois ou para os casos com três ou mais grupos, alguns procedimentos devem ser seguidos para a aplicação da AD. Após definidas e inclusas as variáveis dependentes e independentes, definido os tamanhos dos grupos e verificadas as condições para a utilização da AD, o passo seguinte é a determinação da função discriminante.

Mingoti (2005) explica a construção da função discriminante considerando uma situação básica: duas populações (A e B), com uma variável dependente com dois grupos associados às populações e uma variável independente X (sendo $p=1$, o número de VI). Conhecidas as distribuições de probabilidades da variável explicativa (X) em cada população, supostas normais com médias μ_1 e μ_2 e variâncias iguais (σ^2), para cada possível valor da VI, constrói-se a razão entre as respectivas funções densidades obtendo a chamada razão de verossimilhança entre essas duas populações:

$$\lambda(x) = \frac{\text{função densidade de } x \text{ na população A}}{\text{função densidade de } x \text{ na população B}} = \frac{f_1(x)}{f_2(x)} \quad (1.0)$$

Considerando as duas populações com distribuição normal, tem-se:

$$\lambda(x) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right\}}{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right\}} \quad (1.1)$$

E a regra para classificação seria $\lambda(x) > 1$, $f_1(x)$ é superior àquele obtido em $f_2(x)$, assim, pelo princípio da máxima probabilidade, o indivíduo (x) seria classificado na população A. Se $\lambda(x) < 1$, seria classificado na população B. Para $\lambda(x)=1$, poderia se fazer um sorteio para definir seu grupo ou procurar mais informações (MINGOTI, 2005).

A expressão (1.2) pode assumir outro formato quando se torna em ambos membros o logaritmo neperiano e posteriormente multiplicando-se por -2 obtém-se:

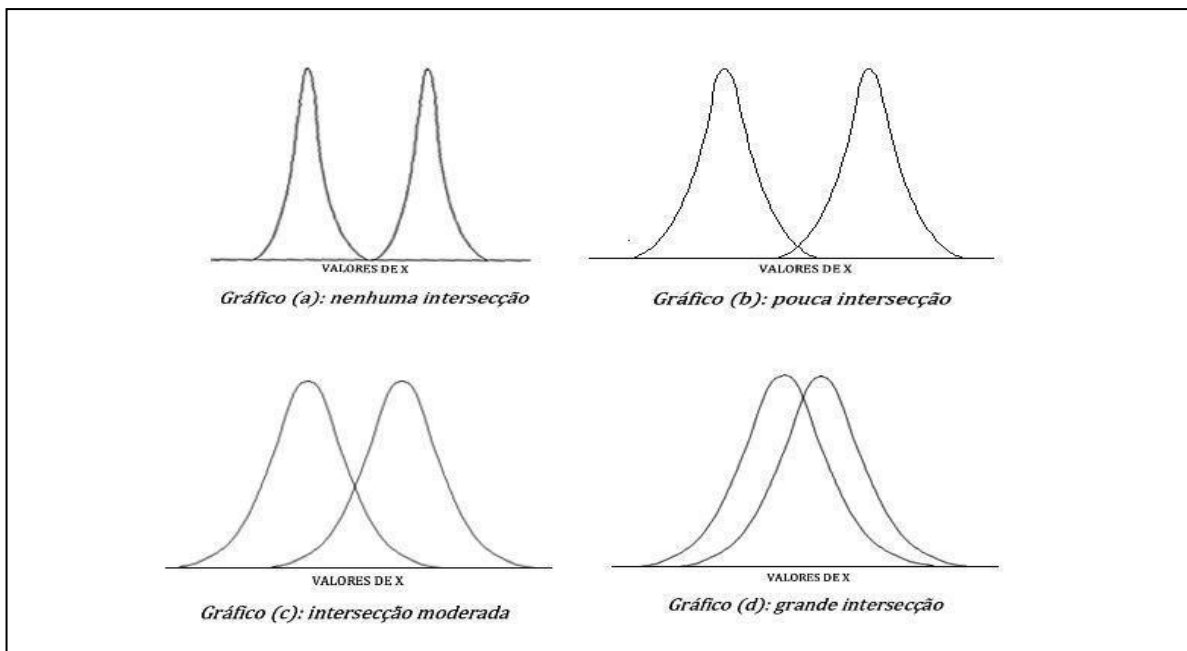
$$-2 \ln(\lambda(x)) = \left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_2}{\sigma}\right)^2 = \frac{1}{\sigma^2} [(x-\mu_1)^2 - (x-\mu_2)^2] \quad (1.2)$$

Nesta forma: quando $-2 \ln(\lambda(x)) < 0 \rightarrow$ escolhe a População A (μ_1)

Essa transformação torna a classificação em termos de distância, isto é, calculam-se as distâncias entre o valor (x) a ser classificado e as médias das duas populações. A menor distância define o grupo a ser escolhido, isto é, o grupo mais “semelhante” (verossimilhança).

A Figura 02 mostra possíveis situações de classificação em duas populações normais de mesma variância.

Figura 02: Populações normais com mesma variabilidade e uma variável discriminante.



Fonte: Mingoti (2005, p.118)

2.1.2.5 Caso de dois grupos com mais de uma variável independente

Um caso mais geral envolve duas populações com variabilidades distintas e um conjunto de “p” variáveis independentes (X_1, X_2, \dots, X_p), tendo um indivíduo em particular, o perfil $x=(x_1, x_2, \dots, x_n)$. Neste caso, as matrizes de variância e covariância são distintas e a regra de classificação seria dada por (MINGOTI, 2005):

$$-2 \ln(\lambda(x)) = -2 \ln \left\{ \frac{(2\pi)^{\frac{p}{2}} \left(|\Sigma_1| \right)^{-1}}{(2\pi)^{\frac{p}{2}} \left(|\Sigma_2| \right)^{-1}} \left[\frac{\exp \left\{ -\frac{1}{2} (x-\mu_2)' \Sigma_2^{-1} (x-\mu_2) \right\}}{\exp \left\{ -\frac{1}{2} (x-\mu_1)' \Sigma_1^{-1} (x-\mu_1) \right\}} \right] \right\} \quad (1.3)$$

$$= [(x-\mu_1)' \Sigma_1^{-1} (x-\mu_1)] - [(x-\mu_2)' \Sigma_2^{-1} (x-\mu_2)] + [\ln |\Sigma_1| - \ln |\Sigma_2|]$$

Onde x é o vetor de dados relativo a um indivíduo (item); μ_1 é o vetor de médias da população A; μ_2 é o vetor de médias da população 2 e Σ_1 e Σ_2 indicam as matrizes inversas de covariâncias (e variâncias). Nesta expressão matricial, o expoente -1 indica matriz inversa e o expoente indica uma matriz transposta.

Quando for possível considerar que as populações têm a mesma estrutura de variabilidade, isto é $\Sigma_1 = \Sigma_2 = \Sigma$, a função simplifica para:

$$-2 \ln(\lambda(x)) = (x-\mu_1)' \Sigma^{-1} (x-\mu_1) - (x-\mu_2)' \Sigma^{-1} (x-\mu_2) \quad (1.4)$$

A qual é equivalente a função conhecida como discriminante de Fisher,

$$fd(x) = (\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2} (\mu_2 - \mu_1)' \Sigma^{-1} (\mu_1 - \mu_2) \quad (1.5)$$

que escrita diretamente em função das VI, torna-se:

$$fd(x) = b_1 X_1 + b_2 X_2 + \dots + b_i X_i \quad b = (b_1, b_2 \dots b_p) \quad (1.6)$$

sendo $b' = (\mu_1 - \mu_2)' \Sigma^{-1}$

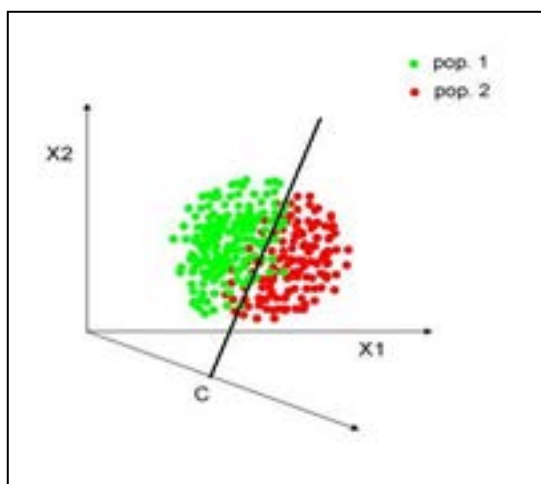
A constante C que separa as regiões dos grupos 1 e 2 é dada por:

$$C = b' \cdot \frac{\mu_1 + \mu_2}{2} \quad (1.7)$$

Daí a regra: se o indivíduo $x = (x_1, x_2, \dots, x_p)$ tem escore dado por $fd(x)$ superior a C , ele é classificado na população A (da média μ_1).

No método de Fisher o objetivo é encontrar coeficientes b_i (b com índice i) “para maximizar a razão da soma dos quadrados de todos os grupos pela soma de quadrados dentro do grupo para $fd(x)$ ”. Esse formato considera que os coeficientes da função de Fisher são equivalentes aos coeficientes de uma regressão múltipla (LATTIN et. al. 2011). A regra de Fisher conduz a mesma regra de classificação vista em (1.6). Nesses dois casos foram consideradas probabilidades, a priori, idênticas nas duas populações (FERREIRA, 2011). A Figura 03 mostra o caso de uma função discriminante de Fisher.

Figura 03: Função discriminante de Fisher.



Fonte: Mingoti (2005, p. 221).

Em certos casos as classificações incorretas podem levar a “custos”, sendo que os custos esperados por incorreção devem ser considerados na expressão da função discriminação (FERREIRA, 2011).

Nas situações reais os parâmetros (μ_1 , μ_2 , σ , e outras) não são conhecidos e devem ser estimados em função dos dados amostrais (FERREIRA, 2011).

As expressões para as estimativas são aquelas vistas na estatística descritiva, tais como média aritmética, variância, combinação de variâncias e correlação ($\bar{X}_1, \bar{X}_2, S^2_1, S^2_2, S^2_c, S_{12}$).

2.1.2.6 Caminhos para a busca da função discriminante

São diversas as terminologias usuais na construção de uma função discriminante. Segundo Pestana e Gageiro apud Fávero et. al. (2009) os coeficientes da função são obtidos através do cálculo dos autovalores, que são dados pela razão da soma de quadrados obtida entre os grupos e a soma dos quadrados dentro dos grupos. Ao se analisar, entende-se que quanto maiores que 1 forem os autovalores, maiores tendem a ser as variações entre os grupos. Esses autovalores (eigenvalues) são determinados em função do quociente entre soma de quadrados entre grupos (SQE) e soma dos quadrados dentro dos grupos (SQD). É evidente que quanto maior for SQE em relação a SQD maior será a possibilidade de diferenças entre os grupos. Portanto, autovalores altos resultam em boas funções discriminantes (FÁVERO et. al. 2009).

De acordo com Hair et. al. (2005, p. 209), compreende-se que a discriminação pode ser obtida quando se determina peso à variável estatística para cada variável, maximizando a variância entre grupos relativa a variância dentro dos grupos. A função, para o caso de “p” VI, é semelhante à expressão de uma análise de regressão, sendo dada por:

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_i X_{nk} \quad (1.8)$$

Onde:

Z_{jk} = escore Z da função discriminante j para o objeto k.

a = coeficiente linear

W_i = peso discriminante para a variável independente i.

X_{nk} = variável independente n para o objeto k.

Fávero et. al. (2009) relatam que esta função é diferente da função discriminante de Fisher; afirmam que a Z_{jk} é utilizada para facilitar a interpretação

dos parâmetros das VI, enquanto que a de Fisher é utilizada com função de classificar observações nos grupos, porém em termos de resultados finais elas conduzem às mesmas conclusões (FERREIRA, 2011).

Pode-se chegar a ela considerando todas as VI (estimação simultânea) ou, passo a passo (*stepwise*), isto é incluindo as VI uma por vez, a partir da mais adequada (melhor discriminatória) e combinando-a com as demais.

2.1.2.7 Testes estatísticos na análise discriminante

No transcorrer da AD vários testes são necessários a fim de se verificar condições iniciais (normalidade) e de aplicabilidade da técnica, bem como definir caminhos durante os passos seguintes. O método estatístico fornece testes para a verificação da normalidade e multivariada dos dados em softwares como o SPSS e SAS (HAIR, et.al. 2005)

2.1.2.7.1 Método *Stepwise*

Segundo Hoking (1976) e Corrar (2012), em um conjunto de variáveis independentes podem existir algumas que pouco influenciam o conjunto de variáveis dependentes. O método *stepwise* é usado, justamente, para manter no modelo as variáveis que mais influenciam na análise, diminuindo desta forma o número de variáveis que compõem a equação.

Neste método observam-se várias interações, adicionando e removendo variáveis, a partir de um critério de seleção. O mais comum é o teste F de Hoking (1976). Durante a aplicação do método *stepwise* temos os seguintes procedimentos:

1. Escolhe-se a variável X_k que possui o maior coeficiente de correlação para entrar no modelo.
2. Uma variável X_p entra no modelo, se o coeficiente de correlação for maior que o anterior, X_p permanece no modelo, caso contrário X_p sai do modelo.
3. X_p sai do modelo e se o coeficiente de correlação for menor que o anterior, X_p fica no modelo, caso contrário, X_p permanece fora do modelo. Este passo é repetido até que não tenha mais X_p para sair do modelo. Terminada esta etapa retorna-se ao

passo 2 e este passo continua até que não tenham mais variáveis para entrar no modelo.

2.1.2.7.2 Testes para inclusão ou retirada de variáveis independentes

Na definição das variáveis independentes é importante considerar apenas aquelas que de fato discriminam, isto é, possibilitam a classificação em um dos grupos.

No caso mais simples com apenas uma VI, um teste “t” de *Student* pode ser utilizado para verificar se os dois grupos apresentam diferenças estatisticamente significativas. Em resumo, compara-se a média da VI no grupo 1 com a média da VI no grupo 2 (LATTIN, et. al. 2011), por meio de um teste de hipóteses. A expressão para t, considerando as variâncias populacionais correspondentes iguais (COSTA NETO, 2001), é dada por:

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{S_c^2}{\bar{n}_1} + \frac{S_c^2}{\bar{n}_2}}} \quad (1.9)$$

Onde S_c^2 é a variância combinada das variâncias das VI em cada grupo. A estatística “t” tem $(n_1 + n_2 - 2)$ graus de liberdade, n_1 e n_2 são os tamanhos dos grupos, sendo H_0 a hipótese de que as médias não apresentam diferenças significativas. Para a VI ser uma discriminante significativa o teste deve rejeitar a hipótese H_0 .

Quando se tem mais de uma variável independente, a comparação se dá com os centroides definidos em cada grupo pelas VI. Nesse caso, o teste estatístico utilizado é o T^2 de Hotelling, dado por:

$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} d' C_w^{-1} d \quad (1.10)$$

Onde d é o vetor de diferenças de médias entre os grupos 1 e 2, $d = (x_2 - x_1)$, d' é a sua transposta e C_w^{-1} é a inversa da matriz de variâncias e covariâncias (LATTIN, et. al. 2011)

Uma transformação permite utilizar a estatística F (Fcal) de Snedecor:

$$F_{cal} = \frac{(n_1+n_2-p-1)}{p(n_1+n_2-2)} T^2 \sim F(p, n_1+n_2-p-1) \quad (1.11)$$

Onde: n_1 e n_2 são as dimensões dos subgrupos e p , é o número de variáveis independentes.

Em termos de notação de testes de hipóteses, H_0 indica que o vetor de médias da população correspondente ao grupo 1 não difere do correspondente para o grupo 2.

Ao nível de significância, fixado (usual 5%) o modelo F, define um ponto crítico dado pelo modelo (Fmod) que separa as regiões de H_0 e H_1 . Comparando o valor calculado (Fcal) em função das amostras com o valor do modelo (Fmod); se $F_{cal} > F_{mod}$, rejeita-se H_0 , indicando que as diferenças entre os grupos, em vista das VI, não se devem ao acaso, isto é, as VI discriminam de forma significativa e podem fazer parte do modelo.

Outros testes, de Wilks (Lambda) e Qui-quadrado, também permitem comparar médias entre os grupos (VD) para cada variável (VI) (FÁVERO et. al. 2009).

O teste de Wilks baseia-se no quociente entre a “soma de quadrados dentro dos grupos - SQD” e a “soma de quadrados total - SQT”, expressões conhecidas na Análise de Variância (ANOVA),

$$\lambda = SQD / SQT \quad (1.12)$$

O modelo de probabilidades é o F de *Snedecor* com p e $(N-p-1)$ graus de liberdades, onde p é o número total de VI, N o tamanho da amostra e o Λ valor de *Lambda*, sendo F para o caso de dois grupos, dado por (FÁVERO, et. al. 2009).

$$F = \left(\frac{1-\Lambda}{\Lambda} \right) \left(\frac{N-p-1}{i} \right) \quad (1.13)$$

Essa expressão não se aplica para o caso de mais de dois grupos, porém, de acordo com MAROCO, apud Fávero et. al. (2009) transformações em (1.13) possibilitam, em certos casos, a realização do teste.

O Lambda (λ) é um valor que pode variar de 0 a 1. De fato, quanto menor mais provável de que as variáveis possuam poder discriminante, pois nesse caso a SQD seria muito inferior a SQT, indicando que a distância entre os centroides dos grupos apresentam diferença significativa.

Em função do Lambda definem-se as variáveis a serem inseridas no modelo. Na prática, as variáveis indicam diferenças significativas a um nível de 5% são escolhidas e, no intervalo de 5% a 10% são reanalisadas. Para diferenças significativas a nível acima de 10% as variáveis são excluídas, pois têm pouco poder de discriminação (Corrar, et. al. 2012).

2.1.2.7.3 Teste para igualdade de matrizes de covariância

Uma condição para aplicação da função de Fisher é que exista a igualdade das matrizes de variâncias e covariâncias (C) dos grupos. Para tal verificação um teste usual é o de Box, onde H_0 indica que não existe diferença entre as matrizes. Para fins práticos a estatística M de Box sofre uma transformação logarítmica (LATTIN, et al., 2011) sendo dada por:

$$B = (1-C) \left\{ \left[\sum_g (n_g - 1) \right] \ln |\mathbf{C}| - \sum [(n_{g-1}) \ln |\mathbf{C}_{(g)}|] \right\}$$

$$C = \left\{ \left[\sum_g \left(\frac{1}{n_g - 1} \right) \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right] \right\}$$
(1.14)

Onde:

p = número de variáveis independentes

n_g = número de observações no grupo g

g = número de grupos

n = tamanho da amostra

C_g = amostra da matriz de covariância interna ao grupo para o grupo g

C = amostra da matriz de covariância interna ao grupo combinada de todos os grupos B tem distribuição Qui-Quadrado com graus de liberdade v, dado por:

$$v = \frac{1}{2} \cdot p (p + 1)(g-1) \quad (1.15)$$

Quando o Qui-Quadrado calculado é maior que o Qui do modelo, rejeita-se H_0 , indicando diferenças significativas entre as matrizes de covariâncias dos grupos.

Quando isto ocorre Lattin et. al. (2011) sugere utilizar a abordagem de *Mahalanobis*, que considera as distâncias de um elemento x aos centroides de cada grupo. Igualando essas distâncias obtém-se a solução para x que graficamente toma o aspecto de uma elipse.

2.1.2.7.4 Testes para avaliar a qualidade do ajuste da função discriminante

Segundo Mingoti (2005), é possível se avaliar a qualidade do ajuste da função discriminante por meio da estatística T^2 de Hotelling, visto em (1.09).

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2) \quad (1.16)$$

Pelo modelo F visto em (1.11) com um nível de significância, digamos de 5%; se o F_{cal} superar o F_{mod} , rejeitamos H_0 , concluindo que o modelo discrimina de forma significativa ($p < 0,05$) os grupos. Caso H_0 fosse aceita, a conclusão é de que esta não seria uma boa função para discriminar e deveria ser reconstruída em função de outras combinações de variáveis independentes.

Outra forma, também simples, consiste em determinar a taxa de acerto e verificar sua significância em função do critério de chance proporcional (LATTIN et. al. 2011).

Ela consiste em classificar as observações usando-se apenas a frequência relativa com que cada grupo ocorre nos dados. Indicando por (p) a probabilidade de se atribuir uma observação (de toda amostra) ao grupo 1 e por (1-p) a probabilidade de atribuir a observação ao grupo 2, pode-se obter o número esperado de classificados se ela fosse feita ao acaso. Na Tabela 02, Lattin et. al. (2011) ilustra a situação. Observe que a soma dos valores da diagonal principal ($np^2 + n(1-p)^2$),

fornece o total de acertos para uma distribuição casual, que resulta em uma taxa de acerto casual (critério da chance proporcional).

Tabela 02: Número esperado de classificados por caso

Real	Grupo 1	Grupo 2	Total
Grupo 1	np^2	$np(1-p)$	Np
Grupo 2	$np(1-p)$	$n(1-p)^2$	$n(1-p)$
Total	Np	$n(1-p)$	N

Fonte: Lattin, et.al. 2011 p.367

Seria essa taxa casual estatisticamente diferente da taxa de acertos devida ao modelo discriminante?

O autor sugere um teste t para comparar esses totais de acertos, considerando aproximação do modelo binomial de probabilidades:

$$t = \frac{A_m - A_c}{S_c} \quad (1.17)$$

Onde,

A_m = acertos devido ao modelo

A_c = acertos devido ao acaso

S_c = desvio padrão da variável “número de acertos ao acaso”

$$S_c = \sqrt{np'(1-p')} \quad (1.18)$$

p' = probabilidade de acertos casuais

Idêntico aos testes anteriores, H_0 representa a igualdade dos acertos e H_1 indica que as variáveis independentes no modelo possibilitam a discriminação de forma significativa.

2.1.2.7.5 Teste de chances

A partir dos resultados da função discriminante definida, os casos são distribuídos em grupos com base em seus escores. Hair et. al. (2005) explica que é possível se avaliar a análise feita com base na distribuição Qui-quadrado do Q Press que verifica o poder discriminatório da matriz de classificação em relação ao modelo aleatório, apontando se a classificação obtida pelo modelo é estatisticamente superior ao da classificação casual. A fórmula é dada por:

$$Q \text{ Press} = \frac{[N-(nk)]^2}{N(K-1)} \quad (1.19)$$

Onde:

N = Tamanho da amostra total

n = Número de classificações corretas

k = quantidade de grupos

A rejeição de H_0 indica que a função discriminante tem bom nível de segregação dos grupos em estudo (CORRAR, et. al. 2012).

2.2 REVISÃO DE LITERATURA

Esta sessão traz uma revisão de artigos relativos à Seguro de automóvel; Sinistralidade e Análise Multivariada; relacionados a perfil de segurados.

2.2.1 Seguros de Automóvel

Quando se pesquisa trabalhos e artigos acadêmicos, cujo tema envolve seguro de automóvel e sinistralidade, tem-se pouca variedade de materiais. A maioria das buscas on-line apontam para textos de palestras, manuais de seguradoras e sites dos próprios órgãos do ramo com dicas e conceitos de seguros.

Gonçalves e Tavarayama (2015) em seu artigo abordaram a importância da viabilidade da utilização do Sistema de Rastreamento e Monitoramento de veículos de passeio e de navegadores GPS para traçar rotas, indicar trajetos e condições da

via em tempo real para o auxílio do condutor. Neste, foram pesquisados corretores de seguradoras do interior de São Paulo e com o resultado de toda a pesquisa verificou-se que sua utilização é importante e satisfatória para a segurança e monitoração dos veículos via satélite, o que em caso de roubo tende a ser uma ferramenta muito utilizada.

Simonovitch, (1997) em sua dissertação trata sobre o risco e análise de risco, definindo e explorando as dimensões dadas pelos estudiosos do assunto e pelos profissionais de seguro, relatando sobre conceitos de risco e técnicas de gestão de risco. Apresenta um estudo de caso de subscrição de apólice de um grande lote de automóveis, derivado de uma situação real, modelando o problema e aplicando o método da simulação como ferramenta de análise de risco para problemas de incerteza na indústria de seguros.

Lugon (2004) desenvolveu um trabalho muito próximo do que pode ser visto nesta dissertação, identificando um modelo para gestão do risco nas propostas de seguro automóvel, baseado em características pessoais do segurado e com o uso da análise discriminante. No geral, foi possível classificar corretamente 54,4% dos segurados pesquisados, apontando a necessidade de novas variáveis no modelo.

Lima (2008) desenvolveu e teve êxito em sua análise de viabilidade financeira baseada na criação de uma célula para a venda de seguros massificados em uma corretora de seguros, utilizando como ferramenta análises de fluxo de caixa, ponto de equilíbrio e de retorno de investimento.

Lugon e Zaquetto, ainda em seu artigo, desenvolveram um modelo de gestão de risco para as propostas de seguros de automóvel a partir de dados fictícios de aspectos sócio econômicos e culturais, visando entender porque indivíduos com características idênticas têm comportamentos tão díspares quando relacionadas à ocorrência de sinistros? E, para tanto, a ferramenta utilizada foi a análise discriminante que resultou em um modelo com baixo percentual de classificação correta, o que indica a necessidade de inclusão de novas variáveis.

2.2.2 Análise Discriminante para a identificação de perfis

Segundo Lugon (2004), é possível se identificar a propensão ao sinistro do segurado a partir das variáveis cadastrais das companhias seguradoras utilizando-se

ferramentas estatísticas. Em seu trabalho utilizou a técnica estatística conhecida como Análise Discriminante. Para a análise foram utilizadas variáveis quantitativas e qualitativas (com o uso de variáveis dummies) que resultaram em 54,9% de casos classificados corretamente.

Para Nóbrega (2010), a AD, com o apoio do software SPSS, pode ser utilizada como instrumento que permite identificar a qual grupo certa observação, produto ou pessoa deve permanecer, uma vez que esta técnica se caracteriza como uma previsão de classificação. Com o uso de dados ilustrativos aplicou a análise em um total de 40 informações e obteve-se uma função para classificar os indivíduos entre 2 grupos.

Os autores Selau e Ribeiro (2009) também utilizaram a AD como uma ferramenta de classificação. Em seu artigo, partindo da necessidade de se identificar o perfil de bons e maus pagadores com critérios objetivos, propuseram uma sistemática para construção de modelos de previsão de risco de crédito usando três ferramentas: análise discriminante, regressão logística e redes neurais.

Em seu artigo Slomski e Mello (2009) utilizam a AD para propor uma função discriminante, com respeito aos indicadores que representam a situação financeira dos estados brasileiros. Depois de selecionados os indicadores relacionados a parte financeira e aplicada a análise discriminante, obteve-se as variáveis que melhor representam a situação financeira dos estados brasileiros, sendo as que fornecem as receitas per capita ($RP / \text{População}$) e a relação entre ativo total menos o passivo total e a receita corrente líquida ($(AT - PT) / RCL$). Os autores concluíram que a situação financeira não se explica por um único indicador, sendo necessário um conjunto de indicadores reunidos estatisticamente.

Comparando este trabalho com outros, verificou-se uma semelhança por utilizarem a Análise Multivariada com o fim de identificar variáveis que definem tipos de perfis.

3. MÉTODO DE PESQUISA

Este capítulo tem como finalidade descrever os procedimentos relativos a pesquisa para a definição das variáveis potenciais da análise discriminante, bem como apresentar os testes estatísticos utilizados e a sequência de procedimentos para se atingir os objetivos.

3.1 PESQUISA

3.1.1 Amostra

Para esse trabalho foram consideradas informações sobre compradores registrados no banco de dados de uma corretora de seguros sediada em Bauru, sendo estas pessoas físicas e com contrato (apólice) ativo no período compreendido entre 01/01/2012 e 01/01/2013, totalizando 2.000 segurados. A escolha desse período decorreu da disponibilidade dos dados, além da necessidade de se analisar um intervalo de doze meses anterior ao ano de 2014, ano da realização da pesquisa.

Em seguida, por meio de um levantamento de dados tipo survey, método de pesquisa que pode ser classificado, segundo Gil (2010), como um instrumento de coleta de dados de um grupo significativo de pessoas acerca do problema estudado para uma abordagem quantitativa, foi extraído da corretora informações de 2.000 segurados que responderam ao questionário da mesma, chamado de Questionário de Avaliação de Risco - QAR.

Esta amostra não pode ser considerada aleatória, ela representa uma quantidade de 1.000 segurados não sinistrados e 1.000 sinistrados, resultando em 2.000 casos fornecidos pela corretora.

3.2 Etapas

O presente trabalho, inicialmente, consistiu de uma pesquisa bibliográfica em livros e artigos científicos encontrados na Web of Science e no Scopus sobre seguros e análise discriminante. Em seguida, foi realizada uma pesquisa de campo,

com o apoio de um questionário (Anexo 1) a fim de se obter informações que pudessem ajudar na definição de variáveis de interesse para o estudo.

Segundo Berto e Nakano (1998), um questionário é o principal instrumento de coleta de dados do método survey e através do Questionário de Avaliação de Risco (QAR) que contém perguntas do tipo: sexo, idade, cargo, quantidade de sinistros, etc..., foram coletadas as variáveis (questões) para a análise multivariada.

Para esta pesquisa foram consultados cinco corretores especialistas, sendo quatro com mais de 15 anos na área de vendas e uma técnica de seguros. Eles foram escolhidos dentre 30 corretores da empresa com mais tempo de serviço da área de vendas de seguros, sendo que a técnica de seguros é uma experiente em cálculos de prêmios. O questionário contém 8 questões e seu principal objetivo foi obter mais informações, não contempladas no Questionário de Avaliação de Risco - QAR (Anexo II), que deveriam ser feitas aos potenciais compradores, de modo a facilitar a identificação do segurado em um dos dois grupos: "Com Sinistro" e "Sem Sinistro". A partir dessa pesquisa foram sugeridas algumas variáveis de interesse, respeitando as limitações do banco de dados fornecido pela corretora.

Em resumo o questionário (Anexo I) abordou:

1. A identificação do entrevistado
2. Confirmação de seu tempo de serviço
3. O que entendia sobre o papel da seguradora
4. O que entendia sobre o papel do corretor de seguros
5. Com quais seguradoras trabalhava
6. Quais as perguntas utilizadas no QAR (Questionário de Avaliação de Risco)
7. Quais perguntas são as mais utilizadas para calcular o risco do seguro
8. Quais perguntas deveriam ser acrescentadas no QAR

Na realidade as seis primeiras questões objetivaram confirmar a experiência dos pesquisados com o tema, sendo que as questões 7 e 8 foram propriamente aquelas que contribuíram para a definição de variáveis de interesse.

3.2.1 Tratamento das potenciais questões

Independente da pesquisa com os especialistas, algumas variáveis de interesse são consagradas neste tipo de análise tais como: sexo, idade, estado civil, região de circulação entre outras. Por exemplo, no caso da idade, variável contínua, alguns autores sugerem, inclusive, a sua categorização a fim de criar mais possibilidades de segregação.

Outras variáveis tais como, quantidade de sinistros, valor do sinistro, mês do sinistro, boletim de ocorrência, com representatividades apenas no grupo dos sinistrados, não são interessantes para a análise discriminante, sendo apenas tratadas de forma descritiva para caracterizar um dos grupos no caso com sinistros.

As variáveis ocupação trabalhista, região de circulação, estado civil, entre outras, caso escolhidas para o modelo, são categorizadas em função das ocorrências na amostra em estudo.

3.2.2 Análises estatísticas dos dados

Os dados da amostra foram dispostos em planilha Excel e organizados em função das necessidades da pesquisa. A estatística descritiva, os testes estatísticos e a análise discriminante foram realizados utilizando-se os softwares Minitab v17 e SPSS v22.

Quanto às variáveis independentes ou explicativas utilizadas no estudo, elegeu-se aquelas que os funcionários pesquisados da corretora sugeriram como sendo importantes, além de outras apresentadas em artigos citados na revisão de literatura, ressaltando as limitações impostas pela empresa fornecedora dos dados.

Na realização dos testes para verificar significância de associações entre variáveis categóricas aplicou-se o teste *Qui-Quadrado* e, para as comparações envolvendo variáveis quantitativas os testes *t* e *Z*, *Box' M*, *U - Wilks' Lambda* e *Q Press* para verificar a significância do poder discriminatório da função. Os testes foram realizados considerando-se nível de significância de 5%. Com base nestes testes foram definidas as variáveis a serem incluídas na Análise Discriminante (AD), isto é, para a definição da função discriminante.

A AD é uma técnica estatística utilizada para identificar quais variáveis diferenciam os grupos e quantas dessas variáveis são necessárias para se

conseguir uma melhor classificação dos indivíduos de uma determinada população nesses grupos (CORRAR et. al. 2012). A Análise Discriminante considerada foi a simples, pois a variável dependente é constituída por apenas dois grupos (Figura 04).

Figura 04: Análise Discriminante



Criou-se então uma variável chamada “grupo” com dois possíveis resultados (0- Não sinistrados e 1- Sinistrados). A empresa forneceu dados de 2.000 segurados, sendo 1.000 não sinistrados e 1.000 sinistrados.

Para a análise discriminante foram removidas as questões: mês do sinistro, valor do sinistro, quantidade de sinistro e boletim de ocorrência, variáveis que só apresentam valores dentro do grupo sinistrado.

Em geral, na AD, as variáveis independentes são contínuas, mas podem ser categorizados (CORRAR, et. al, 2012). Para tanto os dados nominais (algumas questões) foram transformados em variáveis *dummys* (FÁVERO, et. al. 2009), visto que a análise discriminante exige que as variáveis independentes sejam variáveis numéricas. Desta forma, para representar a variável sexo, por exemplo, criou-se duas variáveis *dummys*. Para a variável cargo, onze *dummys*. Para a variável estacionamento criou-se duas *dummys*. Para a variável seguradora criou-se sete *dummys*. Assim as variáveis transformadas em *dummys* passaram a apresentar a seguinte configuração, conforme TABELAS 03 e 04:

Tabela 03: *Dummys* para sexo

Masculino	Feminino
1	0
0	1

Fonte: Adaptado pelo autor

Tabela 04: Dummies para estado Civil

Casado	Divorciado	Solteiro	Viúvo
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Fonte: Adaptado pelo autor

3.2.3 Divisão da amostra total para construção do modelo

Na análise discriminante é usual se dividir a amostra em duas partes, sendo uma para construir o modelo e outra para testar a função obtida (HAIR, et al, 2005). Para tanto no software SPSS foi criado uma nova variável chamada ALEAT, que divide a amostra em duas sub amostras (60% dos dados para a construção do modelo e 40% dos dados para testar a função obtida).

3.2.4 Método para construção do modelo

Após definida a amostra de trabalho (60%), utilizou-se o método *Stepwise* para selecionar quais as variáveis mais influenciam o conjunto de saída, podendo, assim, diminuir o número de variáveis a compor o modelo (FÁVERO et. al. 2009).

O software SPSS em suas saídas fornece as variáveis com maior poder de explicação, e com estas, o software define o modelo gerando os coeficientes da função discriminante.

3.2.5 Verificação do modelo

Para reportar sobre a qualidade do modelo aplica-se o teste Q Press, o qual possibilita verificar a significância do poder discriminatório da matriz de classificação, baseando-se no número das classificações corretas do modelo tanto na amostra de desenvolvimento quanto na de teste. Podendo, assim, verificar se o modelo apresenta um alto poder de separação entre os indivíduos com e sem sinistros.

4. APRESENTAÇÃO DOS RESULTADOS

4.1 Descrição das potenciais variáveis explicativas

Inicialmente procurou-se encontrar variáveis explicativas que aumentassem a chance de se criar uma função discriminante com o maior poder de alocar um indivíduo ao grupo correto (sinistrados e não sinistrados). O objetivo da função discriminante é equalizar as variáveis que melhor possibilitam esta segregação.

Existem variáveis que historicamente já fazem parte deste tipo de estudo, inclusive o próprio QAR que os segurados preenchem fornece questões que são utilizadas pelas seguradoras. Partindo destas questões (variáveis), foram entrevistados quatro corretores e uma técnica em seguros, que sugeriram quais variáveis considerar na AD. Seguem-se essas variáveis.

- Sexo: masculino e feminino.
- Estado civil: situação dos indivíduos em relação ao matrimônio; casados, solteiros, divorciados, viúvos.
- Região de circulação: interpretando a região onde o veículo mais circula. No caso deste trabalho as situações são duas: capital do estado e interior.
- Cargo (ocupação trabalhista): Ocupação e/ou posição que o segurado ocupa dentro de uma empresa ou instituição.
- Potência do veículo: quanto maior for a potência do motor, maior será a sua capacidade de carga e maiores velocidades poderá atingir. Diz-se que um motor é 1.0 ou tem 1 000 cc.
- Idade do segurado (anos)
- Idade do veículo (anos): qual o tempo de uso desde sua fabricação.
- Estacionamento (uso com frequência): as alternativas para este caso foram “Sim” ou “Não”. Indicando se o condutor tem o hábito de guardar seu veículo em estacionamentos com frequência.
- Tempo de seguro (anos): desde quando o segurado contrata seguro de automóvel.
- Bônus: de um ano para o outro as seguradoras utilizam um sistema chamado de bônus-malus, que se resume a diminuir (bônus) o valor pago caso não se

registre sinistros ou agravá-lo (malus), caso se registre um ou mais sinistros. O valor máximo é 10.

- Valor do seguro (Reais): representa o valor pago anualmente pelo segurado à seguradora.
- Seguradora: Instituição responsável por ressarcir o segurado em caso de acidentes, desde que exista um contrato ativo entre as partes (segurado e seguradora) exemplo: Allianz, Bradesco, HDI, CHUBB, Liberty, Yasuda, Tóquio, Sul América, Porto Seguro, Marítima, Mapfre, Itaú, Azul e Alfa.
- Mês da contratação do seguro: mês do ano em que foi contratado.

Partindo das questões sugeridas pelos corretores e das questões usualmente consideradas pela seguradora, o resumo das variáveis definidas para as análises, tanto descritivas quanto para a discriminante, é mostrado na Tabela 05:

Tabela 05: Resumo das potenciais variáveis para as análises

Variáveis de interesse
Sexo;
Idade (anos);
Estado civil;
Região de circulação;
Quantidade de sinistros (0 ou mais que 1)
Cargo (ocupação trabalhista);
Potência do veículo – (cc);
Idade do veículo (anos);
Uso de estacionamento fechado com frequência;
Mês da contratação do seguro;
Tempo de seguro (anos);
Valor do seguro (R\$);
Seguradora;
Bônus.

Fonte: Elaborado pelo autor

É importante destacar que existem questões definidas no QAR, que os corretores não julgaram importantes e que no banco de dados deste estudo não constam. São informações sobre marca, modelo, chassi; se é blindado, valor da blindagem, tipo de residência, quantidade de veículos, distância da residência até o

local do trabalho e tipo de uso do veículo. São dados que em um trabalho futuro poderão ser utilizados.

Com base ainda na questão nº 08 respondida pelos corretores, o QAR hoje é bem completo não havendo necessidade de novas questões. Um dos pesquisados ainda destacou que seria interessante diminuir e não aumentar a quantidade de perguntas feitas ao segurado.

4.2 Características dos segurados da amostra total

Em função da análise descritiva, observa-se que os 2.000 segurados componentes desta amostra são indivíduos que estão na faixa etária de 20 a 80 anos, sendo 64% do sexo masculino. A maioria (76%) é casada, seguida de 16% de solteiros. Em relação ao cargo, 24% exercem atividades em cargos de comando, como administradores, acionistas, diretores, gerentes ou chefias, 22% atuam como analistas, assistentes, consultores. Os segurados de classe média baixa como os técnicos, motoristas, eletricitas, auxiliares administrativos representam cerca de 9% dos dados. Atuando na área jurídica, como magistrados, promotores, advogados encontram-se 7%. Com este mesmo percentual temos os profissionais de humanas, exatas e os aposentados. Abaixo de 7% temos os estagiários (2%), profissionais da área biológica (6%) e do comércio (4%), além de donas do lar (6%). A região de circulação tem como referência cidades ora da capital (49% dos dados), ora do interior (51% dos dados), garantindo uma proporção equilibrada na análise. E, se observados os 2.000 segurados, cerca de 70% utiliza com frequência estacionamentos. Das 14 seguradoras observadas a Allianz e Chubb possuem 70% dos segurados sinistrados. Verificou-se que essas empresas estão mais envolvidas com seguros de prêmios maiores, os quais estão associados a carros mais potentes, que pelos dados têm maior associação com veículos sinistrados.

A tabela 06 mostra as medidas descritivas das variáveis quantitativas.

Tabela 06: Indicadores de variáveis quantitativas

Variáveis	Valor mínimo	Valor Máximo	Mediana	Média	Desvio Padrão
Idade (anos)	20	80	52	52	,976
Valor do Seguro (R\$)	812	56.400	1.746	2.620	3593
Potência (cc)	1.0	4.0	2.0	2.1	,976
Tempo de Seguro (anos)	1	34	14	14	6,32
Idade do veículo (anos)	1	39	7	7	5

Fonte: Elaborada pelo autor

Em relação às variáveis quantitativas, o valor do seguro varia de R\$ 812,00 reais à R\$ 56.400,00 enquanto que o valor do sinistro de R\$ 700,00 a R\$ 6.500,00. A potência dos veículos varia de 1.0 a 4.0 cilindradas. O tempo de seguro, em média, é de 14 anos, enquanto que a idade dos segurados pode chegar aos 80 anos. Quanto a idade dos veículos, metade tem mais de 7 anos de vida, chegando a um máximo 39 anos.

4.3 Seleção das variáveis

Para encontrar as variáveis que mais segregam os grupos (sinistrados e não sinistrados) é fundamental compará-las em função desses grupos. Por exemplo, se no grupo “sinistrado” o tempo médio de seguro não difere estatisticamente desse tempo, no “não sinistrado” esta variável não tem o poder de segregar esses grupos. Para determinar as variáveis a serem utilizadas foram realizados testes de significância. Aplicou-se para as variáveis categóricas o teste Qui-Quadrado e para as não categóricas os testes Z ou t. Com base nestes testes foi definida a variável para o início da construção do modelo da função discriminante.

4.3.1 Testes para as variáveis

A seguir são mostrados os resultados dos testes para verificar a significância da associação entre variáveis nominais ou tratadas como tal (categorizadas), em relação aos grupos. No caso das variáveis contínuas verificou-se a existência de possíveis diferenças significativas quando comparadas nos grupos. As variáveis testadas foram: sexo, idade, estado civil, região de circulação, cargos, potência, uso

de estacionamento, mês da contratação, seguradora, idade do veículo, tempo de seguro, valor do seguro e o Bônus.

- Sexo

A Tabela 07 apresenta resultado do teste Qui-quadrado indicando associação significativa (p -value = 0,006) entre as variáveis sexo e grupo. Apesar de o sexo masculino predominar nos dois grupos ele tem maior percentual no grupo “com sinistro” (33% são mulheres e 67% homens). Nesta tabela os percentuais são mostrados com base nos totais por linhas e colunas. Portanto, a variável sexo é uma potencial participante do modelo discriminante.

Tabela 07: Distribuição dos segurados por grupo em relação ao sexo

Grupos	Feminino	Masculino	Total
0- sem sinistro	389	611	1000
%	38,9	61,1	100,0
%	54,1	47,7	50,0
1- com sinistro	330	670	1000
%	33,0	67,0	100,0
%	45,9	52,3	50,0
Total	719	1281	2000
%	35,9	64,0	100,0
%	100,0	100,0	100,0

Pearson Chi-Square = 7,559; DF = 1; P-Value = 0,006

- Idade do segurado

No caso da variável idade, contínua, o teste t não indicou associação significativa (p -value = 0,344) entre as médias dos grupos com e sem sinistro. Optou-se então por categorizá-la a fim de se obter faixas que indicassem possível segregação. Na Tabela 08, a associação entre as variáveis “faixa etária” e “grupo” indica uma associação significativa (p -value = 0,000), de forma que os indivíduos com sinistros estão concentrados nas classes 1, 4 e 5 (faixas etárias: entre 20 a 30 e de 51 a 70 anos) atingindo um percentual de 53%. Em contrapartida, observando as classes 2, 3 e 6 (faixa entre 31 a 50 e dos 71 a 80 anos) temos um percentual de 58%. Observando os percentuais com base nos totais das colunas, fica mais visível essa diferenciação; dos indivíduos na faixa 1, a maioria pertence a classe com sinistro (57,6%); no outro extremo, faixa 6, os mais velhos são menos frequentes

nesta classe (39,6%). Isto corrobora com a percepção popular de que os mais jovens são mais frequentes nos sinistros.

Tabela 08: Distribuição dos segurados por grupo segundo classes de idades

Grupos	1 (20 a 30 anos)	2 (31 a 40 anos)	3 (41 a 50 anos)	4 (51 a 60 anos)	5 (61 a 70 anos)	6 (71 a 80 anos)	Total
0- sem sinistro	14	174	293	273	136	110	1000
%	1,4	17,4	29,3	27,3	13,6	11,0	100
%	42,4	55,9	52,9	44,1	45,2	60,4	50
1- com sinistro	19	137	261	346	165	72	1000
%	1,9	13,70	26,1	34,6	16,5	7,2	100
%	57,6	44,0	47,1	56,0	54,8	39,6	50
Total	33	311	554	619	301	182	2000
%	1,6	15,5	27,7	30,9	15,0	9,1	100
%	100	100	100	100	100	100	100

Pearson Chi-Square = 26,345; DF = 5; P-Value = 0,000

- Estado Civil

A Tabela 09 apresenta resultado do teste Qui-quadrado rejeitando uma associação significativa (p-value = 0,562) entre o estado civil e grupo. Portanto, esta variável não é uma potencial participante do modelo discriminante. Apesar da tabela apresentar os grupos de divorciados e viúvos com maiores índice de sinistralidade, esta associação não é estatisticamente significativa.

Tabela 09: Distribuição dos segurados por grupo quanto ao estado civil

Grupos	Casado	Divorciado	Solteiro	Viúvo	Total
0- sem sinistro	754	54	167	25	1000
%	75,4	5,4	16,7	2,5	100
%	49,7	47,0	53,2	46,3	50
1- com sinistro	763	61	147	29	1000
%	76,6	6,1	14,7	2,9	100
%	50,3	53,0	46,8	53,7	50
Total	1517	115	314	54	2000
%	75,8	5,7	15,7	2,7	100
%	100	100	100	100	100

Pearson Chi-Square = 2,051; DF = 3; P-Value = 0,562

- Região de circulação

A Tabela 10 apresenta resultado do teste Qui-quadrado indicando associação não significativa ($p\text{-value} = 0,531$) entre a região de circulação e grupo. Portanto esta variável não é uma potencial participante do modelo discriminante. Observados os percentuais na horizontal, por região, tem-se que, mesmo a capital (São Paulo) possuindo uma maior frota de veículos, a diferença de sinistros entre tais regiões não é significativa.

Tabela 10: Distribuição dos segurados por grupo em relação a região

Grupos	Capital	Interior	Total
0- sem sinistro	487	513	1000
%	48,7	51,3	100
%	49,3	50,7	50
1- com sinistro	501	499	1000
%	50,1	50,0	100
%	51,0	49,3	50
Total	988	1012	2000
%	49,4	51,0	100
%	100	100	100

Chi-Square = 0,392; DF = 1; P-Value = 0,531

- Cargo (ocupação trabalhista)

No caso da variável cargo, o Qui-quadrado indicou associação significativa ($p\text{-value} = 0,000$) entre cargos e grupo. Em função do banco de dados apresentar uma série extensa de cargos, procurou-se agrupá-los em categorias que representassem áreas de afinidades com aspectos sociais, financeiros e de formação universitária. A Tabela 11 mostra os cargos agrupados em 11 categorias sendo: **ANA** (analistas, gerentes, supervisores, consultores, encarregados, secretários...), **APO** (aposentados), **BIO** (área de biológicas, como por exemplo: biólogos, dentistas, médicos, nutricionistas, farmacêuticos...), **CC** (classe média baixa, tais como técnicos, motoristas, eletricitas, auxiliares administrativos, escriturários, agentes, técnicos...), **CH** (cargos de chefia, como: acionistas, diretores, presidentes de empresas, proprietários, sócios...), **COM** (setor comercial: vendedores, autônomos...), **EST** (estudantes e estagiários), **EXA** (área de exatas tais como:

engenheiros, contadores, economistas...), **HU** (área de humanas, como, psicólogos, professores, jornalistas...), **JU** (setor jurídico como advogados, delegados...) e **LAR** (donas de casa).

As categorias que mais contribuíram para a associação significativa foram as das classes média-baixa (**CC**), aposentados (**APO**) e Humanas (**HU**) com percentuais superiores a 60% de presença no grupo “sem sinistro”. Por outro lado os segurados que são estudantes (**EST**), ou tem ocupação relacionada a área de exatas (**EXA**), ou ocupações de chefia, direção (**CH**) têm frequências superiores a 60%, porém no grupo dos “com sinistro”.

Tabela 11: Distribuição dos segurados por grupo em relação aos cargos

Grupos	APO	BIO	ANA	CC	CH	COM	EST	EXA	HU	JU	LAR	Total
0- sem sinistro	82	55	240	123	192	41	10	46	85	64	62	1000
%	8,2	5,5	24,0	12,3	19,2	4,1	1,0	4,6	8,5	6,4	6,2	100
%	61,1	47,8	54,2	67,6	39,3	51,2	31,2	35,1	64,4	44,4	52,1	50
1- com sinistro	52	60	204	59	296	39	22	85	47	80	57	1000
%	5,2	6,0	20,3	5,9	29,6	3,9	2,2	8,5	4,7	8,0	5,7	100
%	38,8	52,2	45,8	32,4	60,6	48,7	68,7	64,9	35,6	55,6	47,9	50
Total	134	115	443	182	488	80	32	131	132	144	119	2000
%	6,7	5,7	22,1	9,1	24,4	4,00	1,6	6,5	6,6	7,2	9,5	100
%	100	100	100	100	100	100	100	100	100	100	100	100

Pearson Chi-Square = 83,781; DF = 10; P-Value = 0,000

- Potência do veículo

No caso da variável potência do veículo, o teste t indicou diferença significativa (p-value = 0,000) entre as médias das potências. Na Tabela 12, as potências dos veículos foram agrupadas em 4 categorias de tal modo que todas as variações de 1.0 como 1.1, 1.2 ... constam para a categoria 1, e as variações de 2.0 como 2.1, 2.2... da mesma maneira para a categoria 2 e assim sucessivamente. Se observadas as categorias associadas aos grupos, o percentual com maior sinistralidade está presente entre as potências que variam entre 3.0 a 4.8 cc., o que pressupõe que quanto maior a potência, maior a probabilidade de se ter sinistros:

Tabela 12: Distribuição dos segurados por grupo em relação as classes de potência (cc)

Grupos	1 (1.1, 1.2...)	2 (2.1, 2.2...)	3 (3.1, 3.2...)	4 (4.1, 4.2...)	Total
0- sem sinistro	325	414	174	87	1000
%	32,5	41,4	17,4	8,7	100
%	53,3	51,9	45,1	41,8	50
1- com sinistro	284	383	212	121	1000
%	28,4	38,3	21,2	12,1	100
%	46,6	48,1	54,9	58,2	50
Total	609	797	386	208	2000
%	30,4	39,8	19,3	10,4	100
%	100	100	100	100	100

Pearson Chi-Square = 13,265; DF = 3; P-Value = 0,004

No caso da potência, tanto o teste “t” quanto o teste Qui-Quadrado”, indicaram diferenças entre os grupos.

- Idade do veículo

No caso da variável idade do veículo, contínua, não foi detectada diferença significativa (p-value=0,480) entre as médias dos grupos, as quais ficaram em torno de 7,5 anos. Categorizando as idades dos veículos em faixas de 5 em 5 anos, obtém-se a distribuição da Tabela 13, a qual mostra resultados indicando a não associação significativa (p-value= 0,410) entre as variáveis “idade” e “grupo”. Portanto essa variável não deve ser testada no modelo discriminante.

Tabela 13: Distribuição dos segurados por grupo em relação a idade do veículo

Grupos (anos)	1 (1a5)	2 (6a10)	3 (11a15)	4 (16a20)	5 (21a25)	6 (30a35)	7 (36a40)	Total
0- sem sinistro	408	366	162	49	9	3	3	1000
%	40,8	36,6	16,2	4,9	0,9	0,3	0,3	100
%	52,8	47,8	50,3	43,7	52,9	60,0	60,0	50
1- com sinistro	365	400	160	63	8	2	2	1000
%	36,5	40,0	16,0	6,3	0,8	0,2	0,2	100
%	47,2	52,2	49,7	56,2	47,1	40,0	40,0	50
Total	773	766	322	112	17	5	5	2000
%	38,6	38,3	16,1	5,6	0,8	0,2	0,2	100
%	100	100	100	100	100	100	100	100

Pearson Chi-Square = 6,131; DF = 6; P-Value = 0,410

- Uso de estacionamento fechado com frequência

No caso da variável estacionamento, o teste t indicou diferença significativa (p -value = 0,000) entre as médias dos segurados que utilizam ou não estacionamento e o grupo. Na Tabela 14 observa-se um alto percentual de sinistrados entre os segurados que não usam com frequência estacionamentos. Esta hipótese afirma que tais veículos estão mais expostos a acidentes (sinistros), devido a falta de proteção. Esta conclusão pode ser vista até mesmo entre o grupo de sinistrados que apresenta a proporção similar de indivíduos sinistrados que não utilizam estacionamento.

Tabela 14: Distribuição dos segurados por grupo em relação ao estacionamento

Grupos	Não	Sim	Total
0- sem sinistro	229	771	1000
%	22,9	77,1	100
%	38,4	54,9	50
1- com sinistro	368	632	1000
%	36,8	63,2	100
%	61,6	45,0	50
Total	597	1403	2000
%	29,8	70,1	100
%	100	100	100

Pearson Chi-Square = 46,135; DF = 1; P-Value = 0,000

- Mês da contratação

No caso da variável mês da contratação o teste qui quadrado indicou associação não significativa (p -value = 0,660). Na Tabela 15 pode-se verificar as quantidades de sinistros nos meses do ano. De fato percebe-se uma distribuição de valores próxima de um modelo de distribuição uniforme. Portanto, esta variável não é uma potencial participante do modelo discriminante

Tabela 15: Distribuição dos segurados por grupo em relação ao mês de contratação

Grupos	1	2	3	4	5	6	7	8	9	10	11	12	Total
0- sem sinistro	70	79	83	88	91	80	84	99	88	87	68	82	1000
%	7,0	7,9	8,3	8,8	8,4	9,9	8,8	9,9	8,8	8,7	6,8	8,2	100
%	48,2	44,9	52,2	56,8	52,9	48,5	48,8	52,7	49,2	52,73	46,9	46,4	50
1- com sinistro	75	97	76	67	81	85	88	89	91	78	77	96	1000
%	7,5	9,7	7,6	6,7	8,1	8,5	8,8	8,9	9,1	7,8	7,7	9,6	100
%	51,7	55,1	47,8	43,2	47,1	51,5	51,1	47,3	50,8	47,3	53,1	53,6	50
Total	145	176	159	155	172	165	172	188	179	165	145	179	2000
%	7,2	,8	7,9	7,7	8,6	8,2	8,6	9,4	8,9	8,2	7,2	8,9	100
%	100	100	100	100	100	100	100	100	100	100	100	100	100

Person Chi-Square = 8,583; DF = 11; P-Value = 0,660

- Seguradora

A princípio não se observa motivo para imaginar que a empresa escolhida tenha associação com a variável grupo. Na Tabela 16 é possível visualizar as seguradoras que estão mais associadas com sinistros. Observa-se que a Chubb tem mais clientes sinistrados, porém esta associação pode ser devido ao fato desta seguradora atender clientes de classe social mais alta, que utilizam veículos mais potentes, de alto padrão e estes, sim, estão mais propensos a sinistros (como visto). A categoria Outras inclui (Tóquio, Sul América, Porto Seguro, Marítima, Mapfre, Itaú, Azul e Alfa), e nesta as distribuições são equilibradas. Apesar da significância (p-value = 0,047) próxima do limite usual de 5%, considera-se que a influência dessa variável possa ser explicada por outros fatores, tais como a potência do veículo e/ou ocupação do segurado. Sendo assim não será considerada no modelo.

Tabela 16: Distribuição dos segurados por grupo em relação às seguradoras

Grupos	Allianz	Bradesco	HDI	CHUBB	Liberty	Yasuda	Outras	Total
0- sem sinistro	299	45	32	163	217	153	91	1000
%	29,90	4,50	3,20	16,30	21,70	15,30	9,10	100
%	49,2	47,0	56,0	44,0	51,0	58,0	50,8	50
1- com sinistro	309	50	25	206	210	112	88	1000
%	30,90	5,00	2,50	20,60	21,00	11,20	8,80	100
%	50,8	53,0	44,0	56,0	49,2	42,00	49,2	50
Total	608	95	57	369	427	265	178	2000
%	30,40	4,75	2,85	18,45	21,35	13,25	8,95	100%
%	100	100	100	100	100	100	100	100

Chi-Square = 19,840; DF = 13 (p-value = 0,047)

- Tempo de seguro

No caso da variável tempo de seguro, o teste t indicou diferença significativa ($p\text{-value} = 0,000$) entre as médias dos tempos em cada grupo. Porém, este resultado parece evidente, pois segurados com mais tempo tem maior chance de “sinistros”. Para melhor ilustrar este fato os tempos de seguros foram agrupado em 5 classes. Na Tabela 17 observa-se, ao associar grupo e tempo, na faixa de 1 a 5 anos que a maioria (78,2%) não teve sinistro, porém nas duas últimas faixas (16 a 25 anos) esses percentuais são inferiores a 50% (41,8% e 44,0%). Essa variável também não foi indicada para fazer parte do modelo.

Tabela 17: Distribuição dos segurados por grupo em relação ao tempo de seguro

Grupos (anos)	1 (1 a 5)	2 (6 a 10)	3 (11 a 15)	4 (16 a 20)	5 (21 a 25)	Total
0- sem sinistro	97	324	236	193	150	1000
%	9,70	32,40	23,60	19,30	15,00	100
%	78,2	51,1	57,3	41,8	44,0	100
1- com sinistro	27	310	176	269	218	1000
%	2,70	31,00	17,60	26,90	21,80	100
%	21,8	49,0	42,7	58,2	56,0	100
Total	124	634	412	462	368	2000
%	6,20	31,70	20,60	23,10	18,40	100
%	100	100	100	100	100	100

Pearson Chi-Square = 76,655; DF = 6; P-Value = 0,000

- Valor do seguro

No caso da variável valor do seguro (prêmio), o teste t indicou diferença significativa ($p\text{-value} = 0,000$) entre as médias dos valores dos seguros em cada grupo. O grupo “com sinistro” é aquele que tem, em média, maiores prêmios. Também essa variável foi categorizada, sendo os valores agrupados em 5 classes. Na Tabela 18 é possível observar que quanto menor o poder aquisitivo do segurado, menor é sua frequência de sinistros, tanto que pode-se verificar um número menor de sinistros na classe 1 que corresponde a um valor de prêmio anual entre R\$ 800,00 a R\$ 1.800,00. Com o aumento dos prêmios os percentuais do grupo “sem sinistro” tendem a diminuir.

Tabela 18: Distribuição dos segurados por grupo em relação à classe de valor de seguro

Grupos	1 (800a1800)	2 (1801a3600)	3 (3601a5400)	4 (5401a7200)	5 (7201a12600)	Total
0- sem sinistro	595	294	51	28	32	1000
%	59,50	29,40	5,10	2,80	3,20	100
%	55,4	46,1	39,5	39,4	37,0	50
1- com sinistro	478	343	78	43	58	1000
%	47,8	34,3	7,8	4,3	5,8	100
%	44,5	54,0	60,5	60,6	63,0	50
Total	1073	637	129	71	90	2000
%	53,6	31,8	6,4	3,5	4,5	100
%	100	100	100	100	100	100

Pearson Chi-Square = 33,619; DF = 7; P-Value = 0,000

- Bônus

A variável Bônus, cuja finalidade é apontar quantos anos os segurados tem de seguros sem sinistros, apresenta característica que dificulta associá-la a ocorrência de sinistros, pois a pontuação máxima é de 10 pontos, correspondente a um desconto (%) no prêmio. Porém, ao analisar dois indivíduos com bônus 10 e 14, um deles pode ter zero sinistro no período e o outro até 4. Isto dificulta associar o bônus aos sinistros. De fato, neste caso, não existe diferença estatisticamente significativa ($p\text{-value} = 0,289$) entre os grupos com e sem sinistro quando se

consideram os pontos dos bônus. Isto é, neste formato essa variável não discrimina os grupos.

Construiu-se então a variável Bônus Tempo (BT) como o quociente entre o total de bônus e o tempo de segurado (pontos por ano). Comparando esses valores distribuídos nos grupos, verificou-se diferença estatisticamente significativa ($p\text{-value}=0,000$), sendo que em média o grupo “sem sinistro” apresenta valores superiores (0,776) ao grupo com sinistro (0,664). Essa variável é uma potencial pertencente ao modelo.

4.3.2 Amostra de desenvolvimento e de teste

Considerando a estrutura montada para a análise discriminante, a primeira informação se refere à amostra geral, que foi dividida em duas partes; uma para o desenvolvimento da função discriminante e outra para testá-la. Tal divisão foi feita com base na função ALEAT do SPSS, cujos resultados encontram-se na Tabela 19.

Tabela 19: Dados da amostra subdividida

Tipo de Amostra	Quantidade	Porcentagem
Desenvolvimento	1228	61%
Teste	772	39%
Total	2000	100%

Dos 2.000 elementos disponíveis na amostra, 1.228 (61%) compôs a amostra de desenvolvimento, onde 610 são sinistrados e 618 não sinistrados. Uma questão problemática é quanto a normalidade dos dados, principalmente pelo fato de que algumas variáveis explicativas não são métricas. Hair et. al. (2005), afirmam que “sob certas circunstâncias é possível incluir dados não métricos como variáveis independentes” desde que tratados de forma adequada. Além disso, o tamanho das amostras (610 e 618) em cada grupo é relativamente grande, amenizando possíveis efeitos de não normalidade multivariada.

4.3.3 Variáveis potencias para a construção do modelo

Para o início da construção do modelo, as variáveis definidas em função dos testes realizados em 4.3 são: Sexo, Cargo, Potência, Idade do segurado, Estacionamento, Tempo de seguro, Mês da Contratação, Estado Civil, Valor do Seguro, Região de circulação, Idade do Veículo e Bônus por Tempo.

No procedimento Stepwise no SPSS, um dos métodos utilizados para os testes é o Lambda de Wilks o qual verifica se determinada variável deve ser selecionada ou não para modelo.

O software inicia com a variável mais significativa para a discriminação, no caso Bônus por tempo (BT) (em anos) e na sequência interage com as demais. A Tabela 20 mostra a etapa final com as variáveis representativas para a função discriminante, as quais foram selecionadas considerando os níveis de significância que alcançaram, são elas: BT (bônus por tempo), tempo de seguro, estacionamento, cargos (4,5, 6 e 8) e potência.

Tabela 20: Variáveis consideradas

Etapa	Inseridas	Lambda	Lambda de Wilks	
			Significância	F exato
1	BT (bônus por tempo)	0,931	0,000	90,635
2	Tempo de Seguro (anos)	0,852	0,000	106,064
3	Estacionamento (Sim)	0,835	0,000	80,607
4	Cargo (classe média baixa)	0,822	0,000	66,199
5	Cargo (aposentados)	0,813	0,000	56,319
6	Cargo (área de humanas)	0,805	0,000	49,245
7	Potência (cc)	0,800	0,000	43,696
8	Cargo (analistas, supervisores, gerentes, consultores...)	0,796	0,000	39,089

O teste Box's M, na Tabela 21, confirma que houve violação da premissa da igualdade de matrizes de covariâncias (p-value= 0,000). Segundo Corrar et.al. (2012), seria importante que não houvesse diferença significativa entre as matrizes de covariâncias, porém para grandes amostras espera-se que a influência deste resultado tenha pouca influência no modelo discriminante. O autor afirma ainda que, tal violação possa ocorrer em vista da ausência de normalidade multivariada. No

caso deste trabalho, como o programa continuou gerando os resultados, supõe-se que essas violações “não estão inviabilizando o estudo” (CORRAR et.al. 2012)

Tabela 21: Teste Box' M

Box's M		
M de Box		25,074
Z	Aprox.	25,054
	df1	1
	df2	4508716,022
	Sig.	,000

A tabela 22 ilustra os coeficientes a compor a função discriminante (Escore Z) composta pelos coeficientes não padronizados.

Tabela 22: Coeficientes de função discriminante canônica

Variáveis	Função 1
Potência	-0,178
Tempo de seguro	1,366
BT (bônus por tempo)	10,838
Cargo (analistas, supervisores, gerentes, consultores...)- ANA	0,383
Cargo (aposentados)-APO	1,163
Cargo (classe média baixa)- CC	1,174
Cargo (área de humanas)- HU	0,957
Estacionamento (Sim)	0,624
(Constante)	-12,572

A Função discriminante é dada por:

$$Z = -12,572 - 0,178(\text{potencia}) + 1,366(\text{tempo de seguro}) + 10,838(\text{BT}) + 0,383(\text{cargo}) + 1,163(\text{APO}) + 1,174(\text{CC}) + 0,954(\text{HU}) + 0,624(\text{estacionamento-Sim}) \quad (1.20)$$

A análise desses coeficientes vistos na Tabela 22 sugere que a variável “potência” indica a probabilidade do segurado pertencer ao grupo sinistrado. O “tempo de seguro” indica que quanto menos tempo o segurado tiver, menor poderá ser sua média de sinistros, ao passo que a variável “Bônus por tempo” mostra que quanto maior o numero de bônus em relação ao tempo de seguro, menor a chance dele pertencer ao grupo sinistrado. No caso da variável “cargo” pode-se verificar que

a função trabalhista está ligada ao fator classe social e às áreas do conhecimento (os segurados CC, ANA e HU apresentam um perfil que está mais associado a uma menor ocorrência de sinistros) e, por fim, a variável “estacionamento” aponta que seu uso frequente reduz o número de sinistros. De fato, nos testes realizados em 4.3.1 essas associações indicam resultados estatisticamente significativos.

Apesar da violação a respeito da igualdade de matriz de covariâncias, a leitura da Tabela 23 aponta que esta função tem alta capacidade de discriminar os elementos entre os grupos, visto que o nível de significância está próximo de zero.

Tabela 23: Teste de Wilks' Lambda

Teste de funções	Lambda de Wilks	Qui-quadrado	Df	Sig.
1	,796	279,050	8	,000

No processo de classificação da amostra foram considerados os centroides dos grupos e as respectivas probabilidades para obtenção do ponto de corte (cut-off point) Tabelas 24 e 25.

Tabela 24: Funções para centroides de grupo

Grupo	Função 1
Não Sinistrados	,503
Sinistrados	-,509

Tabela 25: Probabilidades a priori para grupos

Grupo	A priori	Casos utilizados na análise
Não Sinistrados	,500	618
Sinistrados	,500	610
Total	1,000	1228

O ponto de corte serve para a classificação de casos futuros pela função discriminante canônica. O SPSS calcula o ponto de corte considerando custos iguais de erros de classificação, ponderando a relação centroides versus probabilidades. Nesse caso, utilizando o tamanho dos grupos da amostra de desenvolvimento, o ponto de corte identificado foi:

$$\text{Ponto: } \frac{N^2C^1+N^1C^2}{N^1+N^2} = \frac{610(-0,509)+618(0,503)}{618+610} = \frac{(-310,5)+(310,8)}{1228} = 2,44 \quad (1.21)$$

Onde:

N²: quantidade de itens no grupo 1

N¹: quantidade de itens no grupo 0

C²: valor do centroide grupo 1

C¹: valor do centroide grupo 0

Desta forma, focando na amostra de desenvolvimento, o caso em que o valor calculado pela função discriminante canônica for maior que o ponto de corte (2,44), será considerado como um indivíduo sinistrado, caso contrário, como não sinistrado.

Na tabela 26, com o uso da função, é possível alocar na prática quais os indivíduos podem ser direcionados ao grupo dos não sinistrados e quais ao grupo dos sinistrados.

Tabela 26: Perfis de segurados

Não sinistrados					
Pessoa	Potência	Tempo de seguro	BT	Cargo (ANA/APO/CC/HU)	Uso de estacionamento
1222	2	6	0,38	ANA	Não
1908	4	3	0,61	APO	Sim
Sinistrados					
Pessoa	Potência	Tempo de seguro	BT	Cargo (ANA/APO/CC/HU)	Uso de estacionamento
173	2	3	0,91	HU	Sim
1443	4	4	0,53	CC	Não

Onde:

- ❖ Os segurados 1.222 e 1.908 sequencialmente apresentam um valor menor do que o Z crítico 2,44, então classificados no grupo 0:

$$Z = -12,572 - 0,178(2) + 1,366(6) + 10,838(0,38) + 1,174(1) + 0,624(1) = 1,18 \quad (1.22)$$

$$Z = -12,572 - 0,178(4) + 1,366(3) + 10,838(0,61) + 1,163(2) + 0,624(2) = 1,00 \quad (1.23)$$

- ❖ Os segurados 173 e 1.443 sequencialmente apresentam um valor maior do que o Z crítico 2,44, então classificados no grupo 1:

$$Z = -12,572 - 0,178(2) + 1,366(3) + 10,838(0,91) + 1,174(9) + 0,624(2) = 12,85 \quad (1.24)$$

$$Z = -12,572 - 0,178(4) + 1,366(4) + 10,838(0,53) + 1,174(5) + 0,624(1) = 4,42 \quad (1.25)$$

4.3.4 Validação da função discriminante

A partir do ponto de corte, o programa faz a validação da função em relação aos casos das amostras, tanto a de desenvolvimento quanto a de teste. O resultado é visto na Tabela 27:

Tabela 27: Resultados da classificação

Amostras	Grupos	0- Não Sinistrados	1- Sinistrados	Todos Segurados
Desenvolvimento	não sinistrados	353 57%	265 43%	618 100%
	Sinistrados	119 19%	491 81%	610 100%
Teste	não sinistrados	244 59%	158 41%	382 100%
	Sinistrados	81 21%	309 79%	390 100%

Com base na tabela 27, os casos da amostra de desenvolvimento apresentaram 69% de classificações corretas considerando os grupos de não sinistrados e sinistrados. O teste principal é o que se aplica sobre a amostra de teste, reafirmando se a função ou modelo serve para outros perfis de indivíduos. O resultado obtido das classificações corretas alcança um percentual também de 69% de acerto, o que indica que esse modelo tem a capacidade de realizar a classificação de elementos externos a si, isto é, que não contribuíram para este estudo.

Verificando a análise, foi possível ainda validar mais uma vez a função através do Press's Q, que para Hair et. al. (2005) testa o poder de discriminação da matriz

com relação ao número das classificações corretas do modelo. O cálculo foi feito da seguinte forma para a amostra de desenvolvimento:

$$\text{Press's Q} = \frac{[N - (nK)]^2}{N(K - 1)} \quad (1.26)$$

$$\text{Press's Q} = \frac{[1228 - (845.2)]^2}{1228(2-1)} = 173,81$$

N= tamanho da amostra total

n= número de observações corretamente classificadas

K= número de grupos

Desta forma, o Q calculado é 173,81 e considerando um nível de significância de 5%, que resulta em um Q crítico de 3,841, conclui-se que as previsões são significativamente melhores do que um sistema de escolha aleatório. Considerando também o nível de 1% (Q crítico= 6,635), ainda assim esse resultado seria estatisticamente significativo.

5. CONCLUSÃO

Por meio de um rol de informações presentes nos questionários de avaliação de risco (QAR), destacando as características pessoais dos compradores de seguro de automóvel de uma corretora da cidade de Bauru, foi possível identificar as variáveis que possibilitam a segregação dos segurados pesquisados e dos futuros compradores de seguro de automóvel, com determinada probabilidade de acerto nos grupos de sinistrados e não sinistrados. Esta segregação foi possível em função da construção de uma função discriminante simples, a partir de um subconjunto da amostra inicial e outro para o teste desta mesma função.

A separação pode ser feita através de oito variáveis que apresentaram um grau maior de importância: potência, tempo de seguro, bônus tempo, Cargo-ANA, Cargo-APO, Cargo-CC, Cargo-HU e uso de estacionamento. Tais variáveis podem ser consideradas significativamente discriminadoras.

Em meio aos dados, um dos desafios encontrados ao desenvolver este estudo foi o de trabalhar com variáveis qualitativas, que para o modelo da análise discriminante deveriam ser quantitativas. Porém, autores como Corrar et.al. 2012 e Mingoti, 2005 indicam a possibilidade da utilização de variáveis qualitativas. A solução foi converter as qualitativas tal qual a variável Cargo, convertida em uma série de variáveis Dummies.

Esse trabalho apresenta resultados que contrapõe o senso comum, vigente no mercado segurador, no qual os técnicos têm por prática afirmar que o sexo e idade determinam se um indivíduo terá menor ou maior quantidade de sinistros. Esperava-se que os coeficientes destas duas variáveis fossem mais relevantes para o modelo, visto que o maior percentual de sinistros aponta para o sexo masculino.

A amostra de desenvolvimento e de teste apresentaram 69% de classificações corretas entre o grupo de não sinistrados e sinistrados. O que indica que esse modelo tem a capacidade de realizar a classificação dos segurados estudados e de outros potenciais compradores do produto, que não participaram deste estudo.

Para constatar se o uso de dummies influenciou na qualidade da segregação dos segurados, durante o fechamento do trabalho foi rodado novamente o programa considerando apenas as variáveis quantitativas (idade, valor do seguro, idade do veículo, Bonus/tempo e potência) que resultou em uma assertividade de 60% a 61%

de classificações corretas, demonstrando que as variáveis qualitativas, mesmo transformadas em dummies, apoiaram positivamente na segregação dos indivíduos.

Este trabalho tem sua importância e pode servir de apoio às seguradoras que desejam saber a qual categoria seus segurados teriam maior chance de serem alocados, considerando os grupos dos sinistrados e não sinistrados.

Uma sugestão, para garantir uma análise estatisticamente mais aceitável, seria uma reestruturação no formato dos QARs seguindo a Escala Likert, que segundo Oliveira (2001) mede atitudes e comportamentos utilizando opções de resposta que variam de um extremo a outro. Rotular estas escalas, que podem variar de 1 a 5 (por exemplo de ruim para excelente), é um procedimento que facilita a interpretação dos dados.

Outra sugestão, para trabalhos futuros, seria o acréscimo de uma nova variável chamada “omissão segurada”, a qual apontaria a informação de quais segurados solicitaram indenização de sinistro e receberam recusa por parte da seguradora que constatou no QAR informações inválidas ou omitidas pelo segurado, caracterizando o chamado risco moral. Também sugere-se para pesquisas futuras determinar os fatores predeterminantes que influenciam no sinistro e ainda a probabilidade do cliente pertencer ao grupo dos sinistrados mediante a utilização de Análise de Regressão Logística.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBERTI, V.; LEOPOLDI, M. A. P.; MOTTA, M. S.; SARMENTO, C. E.; COSTA, R. C. R.. **Entre a Solidariedade e o Risco: História do Seguro Privado no Brasil**. Rio de Janeiro: Fundação Getúlio Vargas/Funenseg, 1998.

ALVIM, P.. **O contrato de seguro**. 3. ed. Rio de Janeiro: Forense, 1999.

ANDRADE, M. M.. **Introdução à metodologia do trabalho científico**. 4. ed. São Paulo: Editora Atlas S. A., 1999.

BERMAN, B.. **History of Hybrid Vehicles**. HybridCars.com, EUA, 2011

BERNSTEIN, P. L. **Desafio aos deuses: a fascinante história do risco**. Tradução de Ivo Korytowski. Rio de Janeiro: Campus, 1997.

BULCÃO, A. R.; **Seguros: O que você deve fazer**. Paulista Seguros, livro publicado em homenagem ao sexto congresso nacional de corretores de seguros. São Paulo. 1989.

BUSSAB, W.O.; MIAZAK, E.S.; ANDRADE, D.F. Introdução à Análise de Agrupamentos. **9º Simpósio Brasileiro de Probabilidade e Estatística**. São Paulo: IME – USP, 1990.

CASTIGLIONE, L. R.. **SEGUROS: Conceitos e Critérios de Avaliação de Resultados**. São Paulo: Manuais Técnicos de Seguros, 1997.

CHIAPPORI, P.; SALANIÉ B.. Testing for asymmetric information in Insurance Markets. **The Journal of Political Economy**, 108, 1. 2000. 56-78 pp., <http://links.jstor.org/sici?sici=0022-3808%28200002%29108%3A1%3C56%3aTFAll%3E2.0.CO%3B2-J>> Acessado em 19/03/2011

CQCS- **Centro de qualificação do corretor de seguros**. Disponível em: <<http://www.cqcs.com.br/noticia/aumento-de-furtos-e-roubos-em-sorocaba-eleva-preco-seguro-de-carro-em-ate-60/>> acessado em 04/06/14.

COSTANZA M. C. and AFIFI A. A.. Comparison of stopping rules in forward stepwise discriminant analysis, **Journal of the American Statistical Analysis**, 74, 1979, 777-785 pp.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise Multivariada**. 1 ed. São Paulo: Editora ATLAS S.A., 2012.

COSTA NETO, P.L.O., **Estatística**, São Paulo, Edgard Blucher, 2001.

COSTA, P. A.; JUNIOR, P. N. A.; POSSATO, D.S.V.; SALGADO, M. H.; BARRIGA, G. D.C. Aplicação da Análise de Componentes Principais, agrupamento e discriminantes na seleção de características de sinistralidade no Estado de São Paulo. In: **XXXIII Encontro Nacional de Engenharia de Produção; A Gestão dos Processos de Produção e as Parcerias Globais para o Desenvolvimento Sustentável dos Sistemas Produtivos**, 13., 2013, Salvador. **Anais...** Salvador, 2013.

CRASK, M.R., and W.D. PERREAULT, Jr.. "Validation of Discriminant Analysis in Marketing Research," **Journal of Market Research** **14 (February)**, 1977. 60-68 pp.

DENUIT, M., Maréchal, X., Pitrebois, S. & Walhin J.. **Actuarial modelling of claim counts – risk classification, credibility and Bonus-Malus systems**. London: Wiley. 2007. ISBN: 13 978-0-470-02677-9.

DIONNE, G.; Laberge-Nadeau, C.. **Automobile insurance: road safety, newdrivers, risks, insurance fraud and regulation**. Massachusetts: Kluwer Academic Publishers. 1999. ISBN: 978-0-7923-8394-9.

ERSKINE, J. A.; LEENDERS, M. R.; MAUFFETTE-LEENDERS, L. A.. **Teaching with cases**. Ontario: University of Western Ontario. 1981.

ESCOFIER, B.; PAGÈS, J. **Análisis factorialies simples y multiples: objetivo, métodos y interpretación**. Bilbao. Ed. Universidad del Pais Vasco, 1992.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia; SILVA, Fabiana Lopes da; CHAN, Betty Lilian. **Análise de Dados: Modelagem Multivariada para Tomada de Decisões**. Rio de Janeiro: Elsevier, 2009.

FENASEG. **Estatísticas e Projeções para o ano de 2007**. Disponível em: <http://www.fenaseg.org.br>>. Acessado em 06/01/2013.

FERREIRA, D.F. **Estatística multivariada**. Lavras. Editora UFLA, 2008

FERREIRA, D.F. **Estatística multivariada**. Lavras. Editora UFLA, 2011.

FIGUEIREDO, S.. **Contabilidade de Seguros**. São Paulo: Atlas, 1997.

FISHER, L.; VAN NESS J.W., "**Admissible Discriminant Analysis,**" **Journal of the American Statistical Association** 68, 1973, 603-607pp.

FONSECA, M. A. R. **Álgebra linear aplicada a finanças, economia e econometria.** Barueri: Manole, 2003.

FRALEY C.; RAFTERY A.. **Model-Based Clustering, Discriminant Analysis, and Density Estimation.** J. Am. Stat. Assoc., 97 (458): 2002, 611-631pp.

FRANK, R. E., MASSY, W F., MORRISON, D. G.. **Bias in multiple discriminant analysis.** Journal of Marketing Research, 1965, 2, 250-258.

GAMEIRO, A.H.; José Vicente Caixeta Filho. **Administração de seguros no transporte de cargas.** Artigo publicado. Revista RAE-eletrônica. Editora: Fundação Getulio Vargas – Escola de Administração de Empresas de São Paulo. 2002. ISSN 1676-5648.

GALIZA, F. J. S.; DUARTE, L.; e HURTADO, N. H.. **Dicionário de Seguros.** 2ª ed. Rio de Janeiro: Funenseg, 2000.

GILBERTO, F.. **Os Grandes Desafios da Indústria Seguradora.** Lisboa: Lidel, 2010.

GONÇALVES, G. A.; TAVARAYAMA, R.; **Estudo da viabilidade do uso de GPS para o monitoramento de veículos de passeio na região de Orllândia/SP.** ISSUE DOI: 10.3738/1982.2278.605, 2015.

HAIR JUNIOR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise multivariada de dados.** 5. ed. Porto Alegre: Bookman, 2005.

HÄRDLE, W.; SIMAR, L. **Applied multivariate statistical analysis.** 2. ed. Berlin: Springer, 2007.

HOCKING, R. R.. **The Analysis and Selection of Variables in Linear Regression.** Biometrics, v. 32, n. 1, p. 1-49, Washington.1976.

INSURANCE EUROPE. **Como funciona o seguro.** Bélgica.2012

JOHNSON, R.; WICHERN, D. **Applied multivariate statistical analysis.** 3 ed. New Jersey: Prentice Hall, 1992.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Fifth Edition. New Jersey: Prentice Hall, 2002.

KRZANOWSKI, W. J.. "Discrimination and Classification Using both binary and continuous variables." **Journal of the American Statistical Association**, 70. 1975. 782-790 pp.

LACHENBRUCH, P. A. and MICKEY, M. R.. **Estimation of error rates in discriminant analysis**, *Technometrics* 10: 1-11. Li, K.-C (1986).

LATTIN, J.; CARROLL, J. D.; GREEN, P. E. **Análise de Dados Multivariados**. São Paulo: Cengage Learning, 2011.

LARRAMENDI, I. H.; PARDO, J. A.; CASTELO, J. **Manual Básico de Seguros**. Tradução de Maria Armênia de Sousa Vieira. Madri: Mapfre, 1982. 186p.

LEMAIRE, J. Bonus-malus system in automobile insurance. **Massachusetts: Kluwer Academic Publishers**. ISBN: 0-7923-9545-X, 1995.

BRASIL. Lei-nº 4.594, de 29 de dezembro de 1964. Capítulo 1. **Do Corretor de Seguros e da sua Habilitação Profissional**. Disponível em: http://webcache.googleusercontent.com/search?q=cache:cy6J5FNodOgJ:www.planalto.gov.br/ccivil_03/leis/l4594.htm+&cd=1&hl=pt-BR&ct=clnk&gl=br. Acessado em: 29 de jul. de 2014.

LIMA, L. S. **Análise de viabilidade financeira para comercialização de seguros massificados em parceria com empresas associadas ao Sicredi**. Porto Alegre. 77 p. TCC – (Trabalho de conclusão de curso em Administração)- Universidade Federal do Rio Grande do Sul Escola de em Administração), 2010.

LUGON, V.B.P.; **Modelo para gestão do risco nas propostas de seguro de automóvel, com base em características pessoais do segurado, utilizando análise discriminante**. Vitória. 2004. 64f. Dissertação –(Fundação Instituto Capixaba de Pesquisas em Contabilidade, Economia e Finanças (FUCAPE).

LUGON, V.B.P.; ZARQUETTO, H.F.; **Modelo para Gestão do Risco nas Propostas de Seguro de Automóvel, com base no Perfil Sócio-Econômico e Cultural do Segurado, utilizando Análise Discriminante**.

MAGALHÃES, R. A.; **O Mercado de Seguros no Brasil**. Rio de Janeiro: FUNENSEG, 1997.

MAROCO, J.. **Análise Estatística com a utilização do SPSS**. 3ª ed. Silabo. Lisboa. 2007, 822 pp..

MARTINS, G. A. **Estudo de Caso: Uma estratégia de Pesquisa**, São Paulo, Atlas 2006.

MARTINS, João Marcos Brito. **Dicionário de Seguros, Previdência Privada e Capitalização**. Rio de Janeiro: 1ª ed. Forense Universitária, 2005.

MARTINS, João Marcos Brito. **Direito de Seguro**. Rio de Janeiro: 2ª ed. Forense Universitária, 2004.

MATTAR, F.N.. **Pesquisa de marketing** (edição compacta). São Paulo: Atlas. 1996.

MAUSS, M.. **Uma categoria do espírito humano: a noção de pessoa, a noção de 'eu'**. In: Sociologia e antropologia. Trad. Lamberto Puccinelli. São Paulo: EPU, 1974. p. 207-241.

MELLO, G.R. SLOMSKI, V.. **A situação financeira dos estados brasileiros: Uma proposta utilizando análise discriminante**. São Paulo. 2007. Revista de Contabilidade e Organizações, ISSN (Versão impressa): 1982-6486, rco@usp.br, Universidade de São Paulo- Brasil.

MEYER, P.. **Probabilidade - Aplicações à Estatística**, 2ª. ed., Livros Técnicos e Científicos Editora, Rio de Janeiro, 1983.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Ed. UFMG, 2005.

MORRISON, D. F. **Multivariate statistical methods**. New York: McGraw-Hill, 1976.

Mundo Estranho- **Revista Abril**. Disponível em: <http://www.mundoestranho.abril.com.br/materia/como-funciona-o-motor-de-um-carro>. Acessado em: 28 de jun. de 2014.

NÓBREGA D.M.. **Análise discriminante utilizando o software SPSS**. 2010. 54 F, TCC (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologias, Campina Grande, PB, 2010.

OLIVEIRA, C. M. **Contrato de Seguro**. 1ªed. Campinas São Paulo: Editora LZN, 2002.

OLIVEIRA, C. M. **Teoria Geral do Seguro**. Vol. 1. São Paulo: Editora LZN. 2005.

OLIVEIRA, T. M. V.. Escalas de Mensuração de Atitudes: Thurstone, Osgood, Stapel, Likert, Guttman, Alpert. **Revista Administração online** [online]. V.2, n.2. Disponível na internet: <www.fecap.br/admonline/art22/tânia/htm>.ISSN 1517-7912. (2001).

PAUZEIRO, J. C.. **SEGURO: Conceitos, Definições e Princípios**. 2a edição. Rio de Janeiro. VTN Comunicação. 2007.

PEREIRA, C. M. S. **Instituições de Direito Civil**. 11 ed. Rio de Janeiro: Forense, 2004. 604p.

PINHO, F. B., Apostila do Curso de Pós-Graduação de Seguros – PUCAIG, Rocha, J. Guia **Valor Econômico de Seguros- pessoas físicas e bens/SP** Editora Globo, 2003.

RAMOS, C. S.. **Caracterização do acidente de transito e gravidade do trauma: um estudo em vitimas de um hospital de urgência em Natal/RN**. Natal. 2008. 121f. Dissertação – (Universidade Federal do Rio Grande do Norte)

RANGLES, R.H., J.D. BROFITT, J.S. RAMBERG, and R.V. HOGG, "**Discriminant Analysis Based on Ranks**," Journal of the American Statistical Association 73, 1978, pp. 379-384.

REGAZZI, A. J. **Análise multivariada, notas de aula INF 766**, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

RIBEIRO, A. A. C. R.. **Caracterização do perfil das vítimas acidente de transito com motocicleta, na área de abrangência no PSF Boa Esperança no município de Alfenas/MG**. Campos Gerais. 2010. 37f. TCC – (Universidade Federal de Minas Gerais).

RIBEIRO, P. G. **História do Seguro: Um Resumo**. Rio de Janeiro: Funenseg, 1994.

SANTOS, A.. **Dicionário de Seguros**. 2. edição. Rio de Janeiro: Instituto de resseguros do Brasil, 1944.

SANVICENTE, A. Z. **Administração financeira**. 3. ed. São Paulo: Atlas, 1997.

SELAU,L.P.R.; RIBEIRO,J.L.D.; **Uma sistemática para construção e escolha de**

modelos de previsão de risco de crédito. São Carlos. 2009. 16f. – Disponível em: <<http://www.scielo.br/pdf/gp/v16n3/v16n3a07.pdf>> Acessado em: 16 de nov. de 2014.

SHARMA, S.. “Proactive environmental responsiveness: Catalysts for changing organizational paradigms and organizational capabilities.” **BPS, ONE, and SIM. Academy of Management Annual Meetings, Cincinnati, OH, August 1996.**

SILVA, E. G.. **Coletânea Comentada Seguros Previdência Privada Capitalização.** Brasília. CAIXA SEGUROS, 2007.

SILVA, Ernesto V. A. **ABC do Seguro.** São Paulo: Livraria Pioneira, 1984.

SILVA, I. O.. **Curso de Direito do Seguro.** 1º ed. São Paulo: Saraiva, 2008.

SILVEIRA, W.. **O Mutualismo no setor de seguros brasileiro.** 41 F, Monografia – Universidade Federal de Santa Catarina, Florianópolis, 2008.

SIMONOVITCH, M.. **Análise de risco em seguro: um caso prático de uso de ferramentas de simulação de risco na subscrição de apólices de seguro.** São Paulo. 1997. 98f. Dissertação – (Fundação Getúlio Vargas).

SOUZA, S. **Seguros: Contabilidade, Atuária e Auditoria.** 2. ed. rev. e atualizada. São Paulo: Saraiva, 2007.

SUSEP – **Superintendência de Seguros Privados.** Disponível em: <<http://www.susep.gov.br>>. Acessado em: 04 de jun. de 2014.

VIOLA, S. **Noções de Seguros.** Rio de Janeiro: IRB, 1983.

YIN, R. K.. **Applications of case study research.** Thousand Oaks, California: Sage Publications. 1993.

WILLIAMS, C. A.; HEINS, R. M.. **Risk management and insurance.** 5. ed. Nova Iorque: McGrae-Hill, 1985.

ANEXO I – Questionário respondido pelos corretores/técnica de seguros

01 - Identificação do Corretor	
Nome Completo:	
Idade:	
Empresa onde trabalha:	
Susep desde:	
02 - Há quantos anos você trabalha na área de seguros?	
Acima de 5 anos: <input type="checkbox"/>	Acima de 15 anos: <input type="checkbox"/>
Acima de 10 anos: <input type="checkbox"/>	Acima de 20 anos: <input type="checkbox"/>
03 - Qual é o papel da Seguradora?	
04 - Qual é o papel do Corretor de Seguros?	
05 - Com quais seguradoras trabalha?	
06. No questionário de avaliação de riscos quais perguntas são feitas para os clientes:	
() Nome?	() Utilização para ida na faculdade?
() CPF?	() Qual o Estado civil?
() Data de Nascimento?	() Coberturas?
() Qual a Idade?	() Qual o Histórico de sinistralidade?
() Endereço?	() Condutor menor que 24 anos?
() CEP de risco	Outras perguntas:
() Qual Veículo?	_____
() Ano do Veículo?	_____
() Estacionamento fechado ou aberto?	_____
() Utilização para o trabalho?	_____
07. Quais perguntas na sua visão são as mais utilizadas para calcular o risco do seguro? Perguntas que determinam maior possibilidade de um sinistro futuro.	
08. De acordo com sua experiência, quais perguntas deveriam ser acrescentadas no Questionário de Avaliação de Risco?	

Termo de Autorização

Este questionário será utilizado para fins acadêmicos, utilizaremos estes dados para pesquisa e divulgaremos o nome de quem o respondeu para fins de confiabilidade.

Eu, _____,
CPF _____. Declaro estar de acordo com a publicação deste questionário.
Sem mais,

(Assinatura do entrevistado)

Bauru, ____ de _____ de _____.

ANEXO II – Questionário de Avaliação de Risco**Dados do veículo:**

Marca:
 Modelo:
 Ano fabricação/modelo:
 Zero km: Sim () Não() km Previsão de retirada:
 Valor:
 Placa:
 Chassi:ainda não tem
 Veículo Blindado ()Sim ()Não
 Valor da Blindagem:

Dados do segurado:

Nome:
 CPF:
 Data de nascimento:
 Estado Civil:
 RG: Data de expedição: Órgão expedidor:
 Profissão:
 Endereço completo:
 E-mail:

Perfil Principal Condutor:

Nome:
 CPF:
 Data de nascimento:
 Sexo:
 Estado civil:
 Número CNH: CNH: Data 1ª CNH:

Relação do principal condutor com o segurado:

- O próprio Pai/mãe Outros
 Filho(a) Motorista particular
 Cônjuge Diretor/Gerente/Sócio

Existem outros condutores para o veículo?

- Não existem
 Entre 18 e 25 anos
 Acima de 25 anos

CEP de pernoite do veículo:
 CEP da principal região de circulação do veículo:

Tipo de residência

- Casa/sobrado
 Casa em condomínio fechado
 Apartamento ou Flat com porteiro ou portão automático
 Chácara, fazenda ou sítio
 Outros

Possui estacionamento na residência?

- Sim, com portão manual
- Sim, com portão automático ou porteiro
- Não

Possui estacionamento no trabalho?

- Sim
- Não
- Não trabalha ou não utiliza para ir ao trabalho

Possui estacionamento na faculdade/curso?

- Sim
- Não
- Não estuda ou não utiliza para ir a faculdade/curso

Possui quantos veículos?

- ___ Automóveis
- ___ Motocicletas
- ___ Outros

Distância da residência até o local de trabalho:

- Até 10 km Até 20 km Até 30 km
- Até 40 km Acima de 40 km
- Não trabalha ou não utiliza o veículo como meio de transporte ao trabalho

Uso do veículo:

- Somente eventual
- Lazer
- Diário
- Comercial

Quantos quilômetros circula por ano?:

- Até 6.000 km
- De 6.001 km até 12.000 km
- De 12.001 km até 18.000 km
- Acima de 18.000 km

Possui Bônus a ser utilizado?

Possui Bônus a ser utilizado? Sim Não
Classe de Bônus atual:
Seguradora:
Vigência:

Estamos à disposição para quaisquer esclarecimentos.