

1 Introdução

Esta dissertação é o estudo de um problema estatístico de classificação que diz respeito à precificação de seguros de automóveis.

Devido às particularidades deste ramo, a formação dos contratos, a determinação do preço e os serviços prestados apresentam características específicas.

Os elementos essenciais para a formação de um contrato de seguros são: o segurado, o segurador, o objeto de seguro, o prêmio e o risco.

O contratante do seguro é definido como “segurado”, e o “segurador” aquele que oferece o serviço. O “objeto de seguro”, no caso do ramo de automóveis, é representado pelo veículo, e o “prêmio” pode ser entendido como a remuneração paga pelo segurado ao segurador, para que este assuma a responsabilidade de um determinado risco. O prêmio é, portanto, uma nomenclatura específica do ramo para o preço do seguro.

O “risco” é o evento incerto e que provoca danos ao bem segurado. A consumação do risco é chamada de “sinistro”. Existem muitos tipos de sinistros e a indenização dos mesmos está relacionada com a cobertura do risco contratada. No caso do ramo de automóveis, duas são as coberturas do risco comumente contratadas. A “cobertura de casco”, que se refere ao risco de danos ao próprio bem segurado, que dentre outros, podem ser danos parciais, totais e de roubo; e a “cobertura de responsabilidade civil”, que se refere a danos materiais e corporais provocados pelo bem segurado a terceiros.

O “contrato de seguro” representa a relação em que o segurador se obriga a ressarcir o segurado das conseqüências de um evento danoso e incerto, mediante a obrigação do pagamento do prêmio por parte do segurado. A “apólice de seguro” é definida como sendo o ato escrito que constitui a prova formal do contrato de seguro. A validade da apólice, o que pode ser entendido como o período em que o segurador assume a responsabilidade no caso de ocorrência de danos, é chamado de “período de vigência”. Diante do período de vigência é calculada a “exposição ao risco”. A exposição é o período em que a apólice esteve vigente durante um

período de análise estabelecido. A parcela do prêmio referente ao risco é chamada de “*prêmio de risco*”. É comum que este valor seja determinado através da “*taxa de risco*”.

Normalmente o prêmio de risco ou a taxa de risco são calculados para determinadas classes, chamadas “*classes de risco*”, que são formadas através da agregação de contratos que apresentem determinadas características em comum. O conjunto destas classes é chamado de “*classificação de risco*”¹.

Supõe-se que, das definições de prêmio e taxa de risco, derivam os nomes “*precificação*” e “*tarifação*” de seguros.

Não é recente a aplicação de técnicas estatísticas na solução de problemas de precificação de seguros. Um dos primeiros trabalhos que se propõe solucionar o problema de precificação através de modelos estatísticos foi desenvolvido por Mowbray (1914). Nele o autor procura responder a questão de qual deve ser a quantidade de sinistros de uma carteira para que o prêmio de um determinado ano apresentasse *credibilidade total*².

Utilizando um modelo probabilístico que pressupõe a distribuição binomial dos *sinistros*, Mowbray provou que, para um total de 1082 sinistros, o *prêmio puro*³ estimado através da experiência individual de um ano poderia ser generalizado para o ano seguinte.

Embora esta descoberta tenha sido de grande importância para o mercado de seguros da época, não era uma solução para o caso onde o número de sinistros não excedesse 1082.

Whitney (1918) expôs a necessidade de, ao estimar o prêmio, realizar uma ponderação entre a experiência de cada classe de risco e o *risco total*⁴ de uma carteira. Com isso surgiu a fórmula de credibilidade, que estima o prêmio de credibilidade através das estimativas do prêmio por classe e do prêmio total.

A ponderação é realizada através de um fator de credibilidade que assume valores entre 0 e 1. Quando este fator assume valor unitário, a solução para o prêmio é a mesma proposta por Mowbray, denominada de credibilidade total.

¹ Detalhes dos cálculos de algumas destas variáveis serão apresentados no decorrer desta dissertação.

² Credibilidade total, em linhas gerais, seria a capacidade de representar o comportamento da carteira em anos subsequentes. A teoria da credibilidade é um dos alicerces do estudo atuarial.

³ Informalmente, o prêmio puro representa o prêmio de risco acrescido de uma margem estatística.

⁴ Risco total, em linhas gerais, seria o risco de todos os contratos.

Caso contrário, a solução obtida através da média ponderada seria considerada de credibilidade parcial.

Após o trabalho de Whitney, surgiram novas questões a respeito de como estimar o prêmio, por classe risco e para a carteira, e de como estimar o fator de credibilidade.

Neste ínterim, através de um estudo iniciado por Bailey e Simon (1960-1963), surgiu uma nova abordagem para o problema de credibilidade e precificação de seguros. Embora não se proponha a solucionar o problema de estimação apresentado por Whitney, este trabalho foi de grande contribuição para o mercado de seguros, e por muitos anos a técnica desenvolvida pelos autores foi amplamente difundida e aplicada por atuários.

A técnica proposta por Bailey e Simon pretende validar o conjunto de *relatividades*⁵ estimado para as classes de risco através de certos critérios, dentre os quais se inclui a credibilidade.

Anos mais tarde, Buhlmann (1967) demonstra que o problema de tarifação se encaixa perfeitamente no conceito de estimação bayesiana empírica. Isto significa que o parâmetro de risco apresentaria uma distribuição freqüentista, e que seria possível, através da replicação dos dados relacionados a este parâmetro, estimar os parâmetros estruturais desta distribuição. Com isso, Buhlmann propõe um modelo de credibilidade linear que considera que os dados não apresentam tendências, e que possuem o mesmo *volume de risco*⁶. Este modelo pode ser entendido como uma alternativa para estimar o fator de credibilidade, e os parâmetros de risco total e por classe, dentro da composição do modelo proposto por Whitney.

Pouco tempo depois, e em um trabalho conjunto, Buhlmann e Straub (1970) desenvolveram uma extensão do modelo de Buhlmann. Sob esta nova estrutura, os volumes de risco não são mais considerados iguais, assumindo pesos diferentes mediante sua exposição ou importância segurada.

Outros modelos foram desenvolvidos a partir do modelo de Buhlmann-Straub. Como exemplos, podem ser citados: o modelo linear de Hachemeister

⁵ Informalmente, relatividades representam as distâncias entre as classes de risco.

⁶ Em linhas gerais, o volume de risco pode ser entendido como a quantidade de informações de sinistros.

(1975), que generalizou o modelo de Buhlmann-Straub em uma estrutura geral de regressão; e o modelo de regressão não-linear de Vyllder (1976).

Nesta mesma época, em uma seqüência de trabalhos desenvolvidos abordando o tema da credibilidade, Jewell (1974) mostrou a relação entre a família de distribuições exponenciais e a teoria da credibilidade. Provou que o valor esperado da posteriori e o estimador obtido através da fórmula de credibilidade são equivalentes, quando os parâmetros pertencem à família exponencial e as distribuições são conjugadas naturais.

A partir da prova de Jewell surgiram novas soluções para problemas de precificação de seguros, assumindo uma abordagem bayesiana para a teoria de credibilidade. Um exemplo seria o trabalho desenvolvido por Jong e Zehnwirth (1983), que mostra como os modelos de Buhlmann-Straub, Hachemeister e Jewell podem ser estendidos por filtros de Kalman.

Como as distribuições da família exponencial formam a base dos modelos lineares generalizados, é natural que muitos trabalhos, utilizando como principal argumento o trabalho de Jewell, tenham sido desenvolvidos conjugando esta classe de modelos à teoria da credibilidade. Alguns exemplos são os trabalhos de Brown (1988), Brockman e Wright (1992), Nelder e Verral (1997), Holler (1999), Murphy, Brockman e Lee (1999).

Com o desenvolvimento crescente de estudos focados na precificação de seguros, a técnica dos modelos lineares generalizados tornou-se uma das mais conhecidas e aplicadas pelo mercado segurador.

Atualmente, vários são os trabalhos presentes na literatura atuarial que comparam a eficiência de estimação dos modelos lineares generalizados a algoritmos de redes neurais e árvores de decisão. Como exemplos podem ser citados os trabalhos de Francis (2001), Chapados *et alli* (2001), Dugas *et alli* (2003) e Hadidi (2003).

A exaustão na busca por modelos apropriados à estimação do risco deve-se às particularidades de sua distribuição, que apresenta assimetria, um grande número de valores nulos, e um pequeno, porém relevante, número de *outliers*.

Em geral, o objetivo do desenvolvimento destes modelos é prever o risco mediante a classificação adotada.

Este estudo aborda a questão da classificação de risco e seus efeitos sobre a estimação. Suspeita-se que o grande volume de zeros e *outliers* sejam

conseqüências da quantidade de classes que subdividem o risco. Grande parte da bibliografia pesquisada recomenda a redução do número de classes de risco. Normalmente, evidenciam os problemas do excesso de classes na estimação, porém, nenhum estudo que apresentasse soluções matemáticas para este problema foi encontrado.

Agrupar classes de risco implica, primeiramente, na seleção de variáveis capazes de representar as classes de risco, e em um segundo momento, na escolha da técnica de agrupamento a ser aplicada. Acredita-se que técnicas estatísticas convencionais tenderiam a isolar classes em grupos pequenos, incapazes de suportar ao problema da falta de informações; e aglomerar classes distintas em grandes grupos, perdendo assim as características discriminantes.

Sabendo da importância da subjetividade das classes de risco para o mercado segurador, e dos problemas de seleção de variáveis, esta dissertação propõe um método de classificação de risco para seguros de automóveis.

O desenvolvimento do método de classificação proposto por esta dissertação se divide em seis capítulos, sendo este o primeiro. No segundo capítulo serão apresentadas as definições das variáveis a serem utilizadas e algumas particularidades e esclarecimentos sobre a classificação de risco. No terceiro capítulo, o método de classificação de risco será proposto, e as ferramentas estatísticas utilizadas pelo método e as técnicas de avaliação e eficiência da classificação serão apresentadas. O quarto capítulo apresentará a aplicação do método a dados empíricos. No capítulo 5 o método de classificação proposto será avaliado. Nesta avaliação serão considerados a capacidade de atender às expectativas do mercado, e os benefícios trazidos à estimação. Concluindo a dissertação, o sexto capítulo apresentará considerações finais, e sugestões para trabalhos futuros.