

---

**Recebido em**

22 de agosto de 2013.

**Aprovado em**

7 de julho de 2015.

**1. María-Jesús Segovia-Vargas**

Doutora em Economia  
Financeira e Atuarial  
Universidad Complutense  
Madrid  
(Espanha)  
[mjsegovia@ccee.ucm.es]

**2. María-del-Mar Camacho-Miñano**

Doutora em Contabilidade  
Universidad Complutense de  
Madrid (Espanha)  
[marcamacho@ccee.ucm.es]

**3. David Pascual-Ezama**

Doutor em Psicologia  
Universidad Autonoma Madrid  
(Espanha)  
[david.pascual@ccee.ucm.es]

# Seleção dos fatores de risco nas políticas de seguro de automóveis: uma maneira de aprimorar os lucros das companhias de seguro

**María-Jesús Segovia-Vargas.**

**María-del-Mar Camacho-Miñano e**

**David Pascual-Ezama**

*Facultad Ciencias Económicas y Empresariales, Universidad  
Complutense de Madrid, Madrid, Espanha*

Editor responsável: João Maurício Gama Boaventura, Dr.  
Processo de avaliação: Análise duplo-cego

## RESUMO

**Objetivo** – o objetivo deste trabalho é testar a validade do uso de níveis “*bonus-malus*” (BM) para classificar satisfatoriamente os segurados.

**Método** – A fim de alcançar o objetivo proposto e mostrar a evidência empírica, um método de inteligência artificial, a teoria de Rough Set, foi aplicado.

**Resultados** – A evidência empírica mostra que os fatores de risco comuns empregados pela companhia de seguros são boas variáveis explicativas para classificar políticas dos segurados. Além disso, a variável do nível de BM aumenta ligeiramente o poder explicativo dos fatores de risco *a priori*.

**Implicações práticas** – Para aumentar a capacidade de previsão do nível de BM, questionários psicológicos poderiam ser usados para medir as características ocultas dos segurados.

**Contribuições** – A principal contribuição é que a metodologia utilizada para realizar a pesquisa, teoria de Rough Set, não foi ainda aplicada a esse problema.

**Palavras-chave** – companhia de seguros automobilísticos, fatores de risco, sistema de “*bonus- malus*”, teoria de Rough Set, inteligência artificial.



**Revista Brasileira de Gestão  
e Negócios**

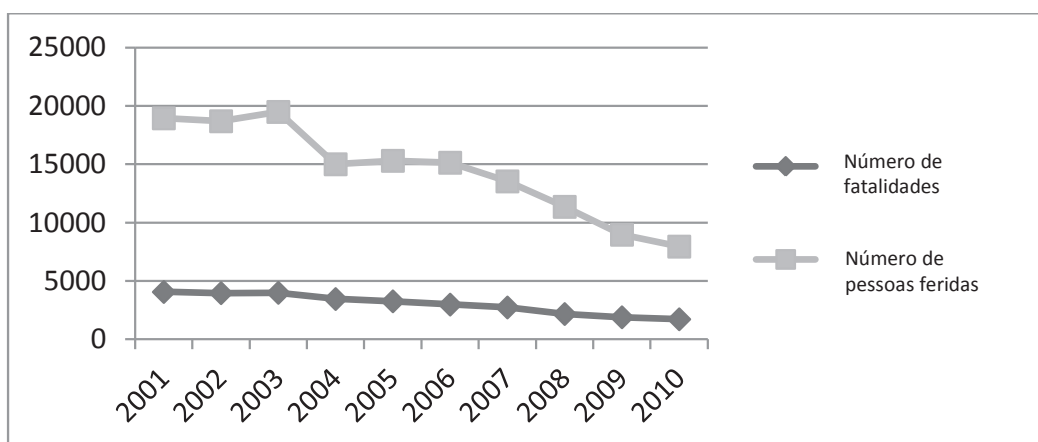
DOI:10.7819/rbgn.v17i57.1741

## I INTRODUÇÃO: ATUAIS PROBLEMAS RELACIONADOS À POLÍTICA AUTOMOBILÍSTICA PARA COMPANHIAS DE SEGUROS ESPANHOLAS

A Espanha é um dos países mais atingidos pela crise financeira que ocorreu na Europa no momento de preparação deste estudo, particularmente em razão da falta de fundos financeiros. Além disso, é um dos países mais relevantes no âmbito da União Europeia (UE), por ser o quinto maior em termos de PIB em 2011, de acordo com o Fundo Monetário Internacional (FMI) (2011). As empresas espanholas estão tentando superar a crise a partir de diferentes abordagens. Para as companhias de seguros, definir riscos envolve a identificação de eventos, suas probabilidades e custos e, embora seja mais fácil para os eventos frequentes, como acidentes rodoviários (Johnson, 2006), também é importante ser eficiente do ponto de vista de gestão, especificamente da eficiência operacional (PwC, 2012). Além de manter os clientes, é necessário gerenciar os riscos corretamente. No entanto, embora o setor de serviços tenha adquirido uma importância crescente em toda a economia do país (Resende & Guimarães, 2012), existem poucas pesquisas dedicadas a serviços de seguros, particularmente àqueles relacionados ao comportamento das partes

envolvidas na oferta ou demanda desses serviços (Silva, 2004).

As companhias de seguros tentam classificar suas políticas de segurados em classes tarifárias homogêneas, atribuindo a mesma remuneração a todas as políticas pertencentes à mesma classe, a fim de cobrar remunerações justas para os motoristas. De fato, “a precisão é, portanto, essencial”, como afirma Arvidsson (2010). Dessa forma, é extremamente importante que a companhia de seguros selecione um conjunto adequado de fatores de risco para prever corretamente as futuras taxas de reivindicação por dois motivos principais. Primeiramente, a competência do mercado de seguros está crescendo, boa parte por causa das ofertas especiais na internet (Segovia-González, Contreras & Mar-Molero, 2009). Em segundo lugar, a taxa de acidentes diminuiu significativamente nos últimos dez anos – especialmente na Espanha, com uma redução de 50% (veja a Figura 1). Por essas circunstâncias, a remuneração das reivindicações poderia efetivamente ser reajustada, uma vez que a probabilidade de altas indenizações diminuiu como resultado. Esses dois reajustes precisam ser limitados pelas companhias de seguros, por razões de viabilidade econômica e para evitar a falência. Conseqüentemente, é necessária uma convergência entre o seguro e a promoção da estabilidade financeira em empresas de seguros.



**FIGURA 1** – Evolução da taxa de acidentes na Espanha (anos: 2001-2010)

Fonte: Spanish National Institute of Statistics (INE) (2015)

Deve-se mencionar aqui que a remuneração paga por uma apólice de seguro de automóvel depende da classe atribuída ao condutor principal. Essa atribuição tem consequências claras para as duas partes afetadas pela escolha do sistema de classificação: a companhia de seguros, em razão dos custos e das receitas incorridos, e o motorista segurado, pela taxa de remuneração. Essa classificação de políticas é baseada na seleção dos chamados “fatores de risco”. Esses fatores são características ou recursos das políticas que ajudam as companhias a prever os valores de reivindicação em determinado período (normalmente um ano). No seguro de automóveis, os fatores são variáveis observáveis relativas ao motorista, ao veículo e ao tráfego. As principais variáveis classificatórias utilizadas pela indústria de seguros são as seguintes: idade, sexo, acidente ou registro de reivindicação do motorista principal, data da carteira de motorista, tipo de veículo e local de residência. Essas variáveis estão correlacionadas com as taxas de reivindicações e, portanto, podem ser úteis para prever as futuras reivindicações. A abordagem comum para selecionar os fatores de risco é baseada em técnicas estatísticas multivariadas, embora essas técnicas ainda deixem uma grande quantidade de heterogeneidade dentro das classes tarifárias. Já existe uma grande quantidade de literatura científica abordando o assunto da classificação de risco dos segurados (Arvidsson, 2010; Denuit, Maréchal, Pitrebois & Walhim, 2007).

Quando, no entanto, os produtos de seguro automobilístico atingem determinada faixa de preço, há muitos fatores importantes que não podem ser considerados *a priori*, por exemplo: rapidez de reflexos ou o comportamento agressivo ao volante. Na verdade, psicólogos têm podido demonstrar que os acidentes em estradas estão relacionados ao comportamento dos motoristas (Aberg & Rimmö, 1998) e às violações de condução (Arvidsson, 2010; Adiante, 2008). Considera-se, então, que essas “características ocultas” são parcialmente reveladas pelo número de reivindicações relatadas pelos segurados

(Pitrebois, Denuit & Walhim, 2006). Assim, a remuneração pode ser reajustada de acordo com o número de reivindicações relatadas pelos segurados. Isso é normalmente feito ao integrar o histórico de reivindicações em um “sistema *bonus-malus*” (BMS). Dessa forma, o BMS é um sistema de classificação de mérito-demérito com duas finalidades: incentivar os segurados a dirigir com mais cuidado e melhor avaliar os riscos individuais para que todos paguem uma remuneração de acordo com seu próprio histórico de frequência de reivindicações (Lemaire, 1988). O BMS é usado em vários países, como a Espanha ou o Brasil. Por conseguinte, as conclusões desta pesquisa são especialmente interessantes para os países que adotaram este sistema de mérito-demérito. No entanto, é relevante, nesta fase, notar que o BMS não é usado em todos os países em razão da maturidade do mercado de seguros e das culturas nacionais (Park, Lemaire & Chua, 2009).

Nesse contexto, é importante salientar que os esquemas do BMS “forçam” os segurados a decidirem se a magnitude de um acidente é suficientemente grande para justificar uma reivindicação, uma vez que fazer uma reivindicação necessariamente implica uma futura perda de desconto. Além disso, os segurados podem ter informações, não observáveis pela seguradora, que preveem o risco ex-pós (Arvidson, 2010). Existem evidências empiricamente demonstradas de que “os motoristas que foram envolvidos em acidentes de trânsito no ano passado se arriscaram mais ao dirigir” (Iversen, 2004). No entanto, enquanto há um debate contínuo sobre os efeitos, problemas e benefícios do BMS, sua utilização pode melhorar a eficiência do mercado (Heras, Vilar & Gil, 2002; Hey, 1985; Richaudeau, 1999).

Quando o BMS é aplicado, a remuneração é calculada ao multiplicar a original por uma porcentagem anexada ao nível do segurado na escala. Isso é conhecido como o coeficiente “*bonus-malus*”. Portanto, o BMS refina a classificação de risco da tarifa *a priori*. Ou seja, um esquema *a posteriori* que usa o BMS pode ser utilizado para

redefinir a taxa *a priori* (Dionne & Ghali, 2005; Pitrebois *et al.*, 2006). Essa atribuição é essencial do ponto de vista financeiro, pois se segurados de alto risco são inadequadamente atribuídos no BMS, a empresa poderia incorrer em um risco elevado de custos. Tal situação poderia prejudicar o futuro da companhia de seguros. Consequentemente, outra variável será considerada (nível BM), com os fatores de risco originais neste estudo.

Tendo tudo isso em mente, o modelo de teste deste trabalho é analisar a previsibilidade de acidentes para todos os fatores com e sem classes de BM, além de fazer uma comparação da previsibilidade de ambos os modelos. Nossa hipótese é a de que o BM pode adicionar informações para melhorar a classificação de políticas de automóveis em classes tarifárias. Além disso, isso será feito para explicar os “fatores ocultos” para uma fixação de preços de seguros precisa. Se o modelo com nível de BM melhora significativamente a explicação da taxa de reivindicações relacionada aos modelos sem nível de BM, os “fatores ocultos” são, consequentemente, suficientemente explicados pela variável de nível de BM.

Este trabalho de pesquisa é dividido em seções. Começa com a seção 2, que mostra o método de Rough Set (RS). Para testar o modelo, o método de RS será aplicado por causa de suas vantagens. Até hoje, não existe um estudo que tenha aplicado essa metodologia para classificar políticas de seguro. Em seguida, a seção 3 descreve os dados e as variáveis. Na seção 4, a metodologia é demonstrada, e a seção 5 discute e apresenta os resultados. Finalmente, as conclusões e propostas serão descritas e destacadas na seção 6.

## 2 METODOLOGIA DE ROUGH SET

A metodologia de RS utilizada para testar os modelos propostos pertence ao domínio da Inteligência Artificial (IA). A IA tem demonstrado um desempenho muito alto ao classificar problemas como os abordados neste estudo. Há, porém,

poucas pesquisas de IA dedicadas ao setor de seguros, embora ele desempenhe um papel crescente e crucial nas economias modernas. Como é o caso com outras metodologias de inteligência artificial, o método de RS tem sido aplicado com êxito para investigar problemas financeiros, como dificuldades financeiras (Ahn, Cho & Kim, 2000; Beynon & Peel, 2001; Dimitras, Slowinski, Susmaga & Zopounidis, 1999; Sanchís, Segovia, Gil, Heras & Vilar, 2007; Slowinski & Zopounidis, 1995), modelagem de viagens baseada em atividades (Witlox & Tindemans, 2004) ou análise de demanda de viagens (Goh & Law, 2003).

No setor financeiro, o bancário recebeu mais atenção dos pesquisadores de IA. As peculiaridades de negócios do setor de seguros, porém, impossibilitam a transferência das descobertas de análises do setor bancário para o de seguros. Portanto, é necessária uma análise específica (D’Arcy, 2005). A maioria dos estudos de IA dedicados ao setor de seguros abordam problemas de insolvência, com resultados satisfatórios (Brockett, Golden, Jang & Yang, 2006; Brockett, Cooper, Golden & Pitaktong, 1994; Díaz, Segovia, Fernández & Pozo, 2005; Kramer, 1997; Martinez de Lejarza Esparducer, 1996; Salcedo Sanz, Fernández Villacañas, Segovia Vargas & Bousoño Calzón, 2005; Salcedo Sanz, Prado Cumplido, Segovia Vargas, Perez Cruz & Bousoño Calzón, 2004; Segovia-Vargas, Salcedo-Sanz & Bousoño-Calzón, 2004). Atualmente, o RS tem sido aplicado no domínio de seguros. Na verdade, Sanchis *et al.* (2007) aplicam o modelo de RS para lidar com a insolvência no setor de seguros a fim de minimizar o risco de falha. Um modelo de 30 regras de decisões foi gerado com alto desempenho em termos de precisão de classificações (80,56%). O modelo de regras mostra, do ponto de vista da solvência, a importância dessas questões: liquidez suficiente, classificação correta, resseguro adequado e a necessidade de ter provisões técnicas suficientes. Por outro lado, Shyng, Wang, Tzeng & Wu (2007) se concentram em descobrir a necessidade do cliente para o mercado de seguros em Taiwan.

Um questionário sobre produtos de seguro foi desenvolvido para compreender as necessidades dos clientes no ano de 2005 com perguntas de única ou múltipla escolha. Os autores aplicaram a teoria de RS para investigar a relação entre um valor único e uma combinação de valores de atributos. Os resultados obtidos com a análise de RS são satisfatórios, pois um teste de ocorrência foi aplicado para verificar a viabilidade das regras de decisão, obtendo um teste com taxa de acerto de 100%. As regras de decisão mostram as seguintes necessidades de seguro dos clientes: o propósito de compra é o financiamento, a remuneração anual média estava em US\$ 938, a idade dos clientes-alvo varia entre 25 e 35 anos e o produto mais comprado é uma mistura de produtos. Além disso, os motivos dados para a não aquisição são falta de interesse e a idade (muito jovens, abaixo de 25 anos).

A seleção do método de RS é baseada não apenas em ser um método de classificação de alto desempenho, mas também em seu caráter explicativo. Se o resultado da classificação for satisfatório, então as conclusões derivadas da metodologia devem ser analisadas.

Essa metodologia tornou-se uma nova maneira valiosa de analisar problemas financeiros, uma vez que apresenta algumas vantagens fundamentais, como o fato de que ela não precisa, normalmente, de variáveis para satisfazer pressupostos. Os métodos estatísticos precisam de variáveis explicativas para satisfazer pressupostos estatísticos, o que pode ser difícil de obter ao se trabalhar com problemas reais. Na seleção de variáveis de dados proposta aqui, existem fatores qualitativos e quantitativos a considerar. Isso pode complicar a análise e os resultados obtidos. Dessa maneira, obtém-se a eliminação de variáveis redundantes, para que o custo do processo de tomada de decisões e o tempo investido pelos tomadores de decisões sejam reduzidos.

De fato, o método de RS ainda não foi aplicado a este problema. Até o momento de

realização desta pesquisa, há apenas um estudo relacionado a procedimentos alternativos para a seleção de fatores de risco. Isso, no entanto, é baseado em métodos de IA *black box* (Bousoño, Heras & Tolmos, 2008). Ou seja, embora os resultados obtidos neste estudo sejam satisfatórios, os métodos de IA aplicados não são tão explicativos quanto os de RS.

A teoria de RS foi inicialmente desenvolvida por Pawlak (1991) nos anos 1980 como uma ferramenta matemática para lidar com a inerente incerteza do processo de tomada de decisões. Apesar de essa teoria ter sido expandida atualmente (Greco, Matarazzo & Solwinski, 1998, 2001), o presente estudo seguirá a abordagem clássica. A teoria de RS envolve um cálculo de partições e está relacionada, portanto, em alguns aspectos, a outras ferramentas que lidam com a incerteza, como a probabilidade estatística ou a teoria de conjuntos difusos. Diferentemente do método de RS, existe uma literatura considerável sobre a teoria de conjuntos difusos na classificação de seguros (Ebanks, Karwowski & Ostaszewski, 1992; Horgby, 1998; Lemaire, 1990; Shapiro, 2005; Wit, 1982; Young, 1996).

A abordagem de RS é de alguma forma diferente da probabilidade estatística e da teoria de conjuntos difusos. Podemos considerar que há três categorias gerais de imprecisão na análise científica. A primeira ocorre quando os eventos são de natureza aleatória. Esse tipo de imprecisão é descrito pela teoria da probabilidade estatística. A segunda ocorre com objetos que podem não pertencer a apenas uma categoria, mas a mais de uma, em diferentes graus. Nesse caso, a imprecisão está associada à forma de difusão na associação do conjunto e é o campo da lógica de difusão. Por fim, a teoria de RS lida com a incerteza produzida quando alguns objetos descritos pelos mesmos dados ou conhecimentos (são, portanto, indiscerníveis) podem ser classificados em diferentes classes (por exemplo, duas companhias com os mesmos valores para algumas variáveis financeiras – elas são *indiscerníveis* – e uma delas vai à

falência e a outra continua em operação), ou seja, não existe apenas uma classificação desses objetos indiscerníveis. Esse fato previne a atribuição precisa a um conjunto. Dessa forma, as classes nas quais os objetos devem ser classificados são imprecisas, mas podem ser aproximadas com conjuntos precisos (McKee, 2000; Nurmi, Kacprzyk & Fedrizzi, 1996).

Essas diferenças mostram uma das muitas vantagens da teoria de RS: não é necessário um agente para estabelecer informações preliminares ou adicionais sobre os dados. Nas outras duas categorias de imprecisão, devem-se atribuir valores numéricos precisos para expressar a imprecisão do conhecimento, como as distribuições de probabilidade em estatísticas ou classe de associação, ou o valor da possibilidade na teoria de conjuntos difusos (Pawlak, Grzymala-Busse, Slowinski & Ziarko, 1995).

O principal conceito dessa abordagem é baseado no pressuposto de que, para cada objeto no universo, há uma correlação com um conhecimento ou dado associado. O conhecimento é tratado nesse contexto como a capacidade de classificar objetos.

A teoria de RS representa o conhecimento sobre os objetos como uma tabela de dados, ou seja, uma *tabela de informações* na qual as linhas são rotuladas por objetos (estados, processos, firmas, pacientes, candidatos...) e as colunas são rotuladas por atributos. As entradas da tabela são valores de atributos. Consequentemente, para cada par objeto-atributo,  $x$ - $q$ , há um valor conhecido chamado de *descriptor*,  $f(x, q)$ . No problema que será analisado, a tabela de informações é composta pelas políticas e os fatores de risco, ou seja, os objetos serão cada segurado e as colunas serão os fatores de risco utilizados (consulte a Tabela 1). Portanto, o descriptor será o valor do fator de risco para cada segurado.

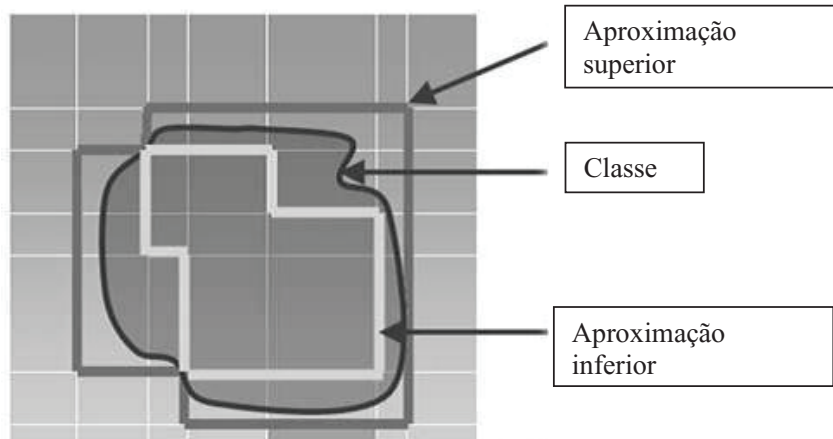
Ocasionalmente, os objetos descritos por alguns dados ou conhecimento são indiscerníveis na visão de tal conhecimento. A relação de indiscernibilidade leva à base matemática da teoria

de RS. Intuitivamente, um RS é uma coleção de objetos que, no geral, não podem ser caracterizados precisamente em termos dos valores de um conjunto de atributos. Em problemas ou bancos de dados reais, a ocorrência de inconsistências nas classificações normalmente aparece. No caso do estudo, há duas classes no banco de dados (motoristas com e sem acidentes). Se um *bom* motorista (sem acidentes) tem os mesmos valores para todos os atributos (fatores de risco) que um *mau* motorista, é difícil classificá-los corretamente em suas classes correspondentes. Matematicamente, a *relação de indiscernibilidade* pode ser expressa em termos de descrições, ou seja, dois objetos,  $x$  e  $y$ , têm todas as suas descrições na tabela com os mesmos valores, ou seja, se e somente se  $f(x, q) = f(y, q)$ .

Existem diversas maneiras para encontrar a solução: a primeira consiste em aumentar a informação (por exemplo, considerar mais atributos), o que, às vezes, não é fácil ou possível. Outra possibilidade é eliminar essas *inconsistências*, o que não é uma maneira correta, pois pelo menos uma informação será perdida. Finalmente, outra maneira é lidar com essas inconsistências ao incorporá-las à análise (o caso de RS). A metodologia de RS incorpora essas inconsistências, criando algumas aproximações para as classes de decisão. A aproximação inferior de uma classe ou categoria consiste em todos os objetos que certamente pertencem a essa classe e podem ser corretamente classificados nessa categoria, empregando o conjunto de atributos (no caso do estudo, os fatores de risco). A aproximação superior de uma classe contém objetos que possivelmente pertencem a essa classe e podem ser classificados nessa categoria usando o conjunto de atributos. A diferença entre a aproximação inferior e superior, se existir, é chamada de limite ou região duvidosa: o conjunto de elementos que não podem ser certamente classificados em uma classe, levando em consideração o conjunto de atributos. Usando a aproximação inferior e a superior, essas classes que não podem ser expres-

sadas exatamente (ou seja, uma região duvidosa) podem ser definidas precisamente utilizando os atributos disponíveis.

A Figura 2 representa graficamente a aproximação superior, a aproximação inferior e a região de fronteira de uma classe ou categoria.



**Figura 2** – Aproximações da teoria de Rough Set

Fonte: Adaptado de “Rough Set, Their Extensions and Applications”, de Q. Shen

e R. Jensen, *Journal of Automation and Computing*, 4, p. 218.

Um problema fundamental da abordagem de Rough Set é identificar as dependências entre os atributos em um banco de dados, uma vez que isso permite a redução de um conjunto de atributos ao remover aqueles que não são essenciais para caracterizar o conhecimento. Esse problema será referenciado como redução de conhecimento ou, em termos mais gerais, um problema de seleção de recursos. O problema de seleção de recursos implicaria a possibilidade de classificar corretamente os objetos sem usar a totalidade dos atributos que foram considerados originalmente. Essa é uma questão muito útil, pois permite que um tomador de decisões classifique, com foco nas variáveis relevantes, o que reduz tempo, esforço e custo em um processo de tomada de decisão. Na medicina, por exemplo, esse fato implicaria diagnosticar um paciente mais rapidamente se algumas provas (especialmente as mais dolorosas ou que consomem tempo) pudessem ser evitadas caso a experiência demonstre que elas não oferecem informações adicionais para o diagnóstico de uma doença. No problema considerado,

os fatores de risco poderiam ser reduzidos sem classificar incorretamente as políticas.

Na teoria de RS, há diversos modelos para reduzir o número de atributos. Um dos mais populares é o sugerido por Skowron e Rauszer (1992). Ele propôs representar a tabela de informações em uma matriz de diferenciação. É uma matriz simétrica na qual as linhas e colunas são os objetos (políticas neste caso, por exemplo  $x_i, y$  e  $x_j$ ). Cada entrada da tabela ( $c_{ij}$ ) representa o atributo ou conjunto de todos os atributos (fatores de risco, neste caso) que podem diferenciar  $x_i$  de  $x_j$ . Comparando cada objeto com o restante em termos de atributos, é possível calcular essa matriz e o núcleo, e os redutos serão obtidos. Um reduto é um subconjunto mínimo de atributos que fornecem a mesma classificação que um conjunto de todos os atributos. Se houver mais de um reduto, a intersecção de todos eles é chamada de núcleo e é o conjunto dos atributos mais relevantes da tabela. Se nenhum dos atributos for redundante, é impossível obter um reduto e, portanto, será necessário utilizar todas as variáveis. Se houver pelo menos um reduto, porém, é possível eliminar todos os atributos que não pertencem a ele,

pois eles são redundantes, ou seja, não fornecem informações adicionais.

Uma vez que a eliminação das variáveis redundantes é obtida, o modelo pode, a partir daí, se desenvolver para o formato das regras de decisão. Além disso, essa técnica é explicativa e gera regras de decisão sem o seguinte formato: “se há condições, então há decisões”. Isso significa quais decisões (ações, classificações) devem ser tomadas quando algumas condições são satisfeitas. O número de objetos que satisfazem a parte de condições da regra é chamado de ponto forte da regra. As regras obtidas normalmente não precisam ser interpretadas por um especialista, uma vez que são facilmente compreendidas pelo usuário ou tomador de decisões. Diversos algoritmos podem desenvolver regras com base na teoria de RS. Bazan, Nguyen, Nguyen, Synak e Wróblewski (2000) desenvolveram a regra implementada no software empregado na parte empírica deste artigo.

Em resumo, o resultado mais importante da abordagem de RS é a geração de regras de de-

cisão, pois elas podem ser utilizadas para atribuir novos objetos a uma classe ao corresponder à parte da condição de uma regra de decisão à descrição do objeto. Portanto, as regras podem ser utilizadas para o suporte à decisão.

### 3 SELEÇÃO DE DADOS E VARIÁVEIS

Foi empregada uma amostra real de 5.500 políticas automobilísticas espanholas observadas durante o ano de 2005. Esses dados foram fornecidos por uma grande companhia de seguros de automóveis da Espanha, embora eles normalmente não sejam disponibilizados por causa da legislação de privacidade e questões de confidencialidade industrial. Os fatores de risco (variáveis) empregados pela companhia são os 14 a seguir (Tabela 1), que exibem uma mistura de variáveis qualitativas e quantitativas e geralmente são aceitos no setor de seguros:

**TABELA 1** – Definições das variáveis utilizadas

Fatores de risco	Códigos computacionais	Explicação
Tipo de veículo	TV	Esta variável leva em consideração seis valores, incluindo carro, van, veículo 4x4, carro misto, van mista e veículo adaptado.
Uso	USO	Uso ao qual o veículo é dedicado. São considerados 12 valores: particular, táxi, veículos de emergência, carros escolares, empresa de carros, empresa de carros temporários, exibição, distribuição, transporte, aluguel com e sem motorista e uso agrário
CV	PV	Potência do veículo (cavalos de potência, HPA)
Privado	PvP	Uso do veículo: privado ou público
Tara	TAR	Tara (peso)
Assentos	NA	Número de assentos do veículo. Considera os seguintes valores: 2, 3, 4, 5, 6, 7, 8, 9.
Âmbito	AC	Área de circulação do veículo. Esta variável considera seis valores: internacional, nacional, regional, interurbano, urbano e rural.
Idade do veículo	IV	Idade do veículo em anos
Idade do segurado	IS	Idade do segurado em anos
Carteira de motorista	EXP	Anos de validade da carteira de motorista
Sexo	SEX	Masculino (M) ou feminino (F)
Região	REG	A área geográfica do segurado onde o veículo está registrado. Foram incluídas todas as regiões espanholas autônomas (Andaluzia, Ceuta, Castilla León, Castilla La Mancha, Cantábria, Baleares, Madri, País Basco, Múrcia, Extremadura, Comunidade Valenciana, Navarra, Aragão, Catalunha, Astúrias, Galícia, Melilla, Canárias, Rioja) e algumas grandes cidades espanholas, como Valência, Barcelona e Sevilha. Esta variável inclui 22 valores.
Combustível	DoG	Diesel (D) ou gasolina (G)
<i>Bonus-Malus</i>	BON	Níveis “ <i>bonus-malus</i> ” nos quais os segurados são classificados pela companhia. São 14 níveis. Um nível mais baixo indica uma remuneração inferior (melhor bônus) – portanto, o nível 1 é o ponto de partida.

Fonte: Dados da companhia.



De acordo com a Tabela 2, a amostra é descrita. A maior parte das políticas do banco de dados pertence a homens (74,3%). Das políticas automobilísticas, 80% são de carros e, especificamente, o uso principal é quase 100% privado. Os veículos da amostra são movidos a

gasolina em quase 6 de cada grupo de 10 casos. A zona de uso do carro é urbana em 74,2% dos casos, ao passo que 22,3% é rural. A amostra é concentrada em duas regiões: Andaluzia (28,3%) e Madri (10,5%).

**TABELA 2** – Frequências das variáveis relacionadas à seguradora, ao carro e à zona de circulação

Códigos	Categorias	Frequência	Percentual	Percentual válido	Percentual acumulativo	
<b>Sexo</b>	Masculino	4084	74,3	74,3	74,3	
	Feminino	1416	25,7	25,7	100,0	
<b>TV</b>	Carro	4394	79,9	79,9	79,9	
	Van	2	0	0	79,9	
	Veículo 4x4	394	7,2	7,2	87,1	
	Carro misto	347	6,3	6,3	93,4	
	Van mista	85	1,5	1,5	94,9	
	Veículo adaptado	278	5,1	5,1	100,0	
	<b>PvP</b>	Público	47	0,9	0,9	0,9
	Privado	5453	99,1	99,1	100,0	
<b>Uso</b>	Particular	5358	97,4	97,4	97,4	
	Táxi	18	0,3	0,3	97,7	
	Veículo de emergência	1	0	0	97,8	
	Carro escolar	13	0,2	0,2	98,0	
	Empresa de automóveis	48	0,9	0,9	98,9	
	Empresa de automóveis temporários	2	0	0	98,9	
	Exibição	2	0	0	98,9	
	Distribuição	10	0,2	0,2	99,1	
	Transporte	1	0	0	99,1	
	Aluguel com motorista	29	0,5	0,5	99,7	
	Aluguel sem motorista	17	0,3	0,3	100,0	
	Uso agrário	1	0	0	100,0	
	<b>DoG</b>	Diesel	2271	41,3	41,3	41,3
		Gasolina	3229	58,7	58,7	100,0
	<b>AC</b>	Rural	1224	22,3	22,3	22,3
Urbana		4081	74,2	74,2	96,5	
Interurbana		125	2,3	2,3	98,7	
Regional		28	0,5	0,5	99,2	
Nacional		2	0	0	99,3	
<b>REG</b>	Internacional	40	0,7	0,7	100,0	
	Melilla	17	0,3	0,3	0,3	
	Ceuta	18	0,3	0,3	0,6	
	Rioja	38	0,7	0,7	1,3	
	Cantábria	106	1,9	1,9	3,2	
	Navarra	134	2,4	2,4	5,6	
	Astúrias	151	2,7	2,7	8,3	
	Baleares	145	2,6	2,6	10,9	
	Extremadura	446	8,1	8,1	19,0	
	Aragão	203	3,7	3,7	22,7	
	Múrcia	175	3,2	3,2	25,9	
	C. Mancha	336	6,1	6,1	32,0	
	Canárias	2	0	0	32,0	
	P. Basco	218	4,0	4,0	36,0	
	C. León	298	5,4	5,4	41,4	
	Galícia	272	4,9	4,9	46,3	
	Valência	346	6,3	6,3	52,6	
	Madri	575	10,5	10,5	63,1	
	Barcelona	285	5,2	5,2	68,3	
	Catalunha	176	3,2	3,2	71,5	
Sevilha	373	6,8	6,8	78,3		
Andaluzia	1186	21,6	21,6	100,0		
	Total	5500	100,0	100,0	100,0	

Outras informações (Tabela 3) da amostra são: a potência média do automóvel é de 91,36 cavalos, o peso médio dos veículos é de 1.198 kg, número médio de assentos é de aproximadamente

5 e a idade média do veículo é de 6,5 anos. A idade média dos segurados é acima de 44 anos e a experiência média da licença para dirigir é de mais de 20 anos.

**TABELA 3** – Estatísticas descritivas das principais variáveis da amostra

Reivindicação		PV	TAR	NoA	IV	IS	VAL	BON
Não acidente	N	2853	2853	2853	2853	2853	2853	2853
	Média	90,4048	1166,2471	4,9930	6,9113	44,8591	20,7774	3,0277
	Desvio padrão	28,66927	292,12331	58973	21,10206	14,11033	11,58152	2,59847
Acidente	N	2647	2647	2647	2647	2647	2647	2647
	Média	92,3993	1232,9686	4,8515	6,0650	44,0147	20,1980	3,2116
	Desvio padrão	28,32268	359,00620	80968	5,01882	12,79749	10,82368	3,14247
<b>Total</b>	N	5500	5500	5500	5500	5500	5500	5500
	Média	91,3647	1198,3584	4,9249	6,5040	44,4527	20,4985	3,1162
	Desvio padrão	28,51782	327,69976	70768	15,59640	13,49982	11,22590	2,87437

Essas políticas são atribuídas a duas classes: acidente = A ou não acidente = N\_A. É importante observar que os segurados são atribuídos a essas duas classes levando em consideração as reivindicações relatadas (ou seja, quando um acidente é relatado, o segurado é reclassificado para a classe “acidente”), porém não seus custos. Isso acontece porque o BMS penaliza somente o número de reivindicações no mundo todo (com pouquíssimas exceções na Coreia) (Lemaire, 1995). Dessa forma, aparentemente a classe de acidentes seria muito heterogênea, mas ela é qualificada considerando os 14 níveis de BM utilizados pela companhia.

#### 4 METODOLOGIA DE PESQUISA

O problema financeiro abordado é um problema de classificação, então os *novos* segurados descritos por um conjunto de fatores de risco são atribuídos a uma categoria (acidente ou não acidente). A fim de alcançar o objetivo, dois modelos foram desenvolvidos: primeiramente, um sem o nível de BM e, em seguida, outro com nível de BM. Se a precisão da classificação (porcentagem dos segurados corretamente classificados) no primeiro modelo é maior do que a

do segundo modelo, então o nível de BM é uma variável redundante. Se, pelo contrário, a precisão da classificação no segundo modelo é superior ao primeiro, o nível de BM inclui os “fatores ocultos” para a elaboração de preços precisa do seguro. Dependendo das diferenças entre os dois modelos (com e sem nível de BM), o poder explicativo da variável do nível de BM é apresentado. Os dois modelos de RS são obtidos para explicar a variável dependente (reivindicações) sem e com a variável do nível de BM.

Se um modelo de classificação fosse desenvolvido e testado com a amostra inteira, os resultados obtidos poderiam ser condicionados. Portanto, para evitar que isso acontecesse, duas amostras aleatórias foram selecionadas: um conjunto de treinamento para desenvolver o modelo (4.400 políticas; ou 80% da amostra inteira) e uma amostra de teste para validar as regras (1.100 políticas; ou 20% da amostra total). O software utilizado para realizar a análise permite a divisão aleatória da tabela em duas subtabelas. É necessário, no entanto, especificar o *fator de divisão* para determinar o tamanho da primeira subtabela, e a segunda será um complemento da primeira. O fator de divisão foi definido como 0,8.

A análise de Rough Set foi realizada utilizando o RSES2.<sup>1</sup> Esse software segue passo a passo

todos os conceitos previamente explicados sobre a teoria de RS. Antes de executar o software, as variáveis contínuas (potência do veículo, tara, idade do veículo, idade do motorista e experiência da carteira de motorista) foram gravadas em termos qualitativos. Essa gravação foi feita dividindo o domínio original em subintervalos. Embora não seja imposto pela teoria de RS, isso é muito útil para retirar as conclusões gerais das regras de decisão ou para interpretá-las (Dimitras *et al.*, 1999). A companhia, com base em suas análises internas, estabeleceu alguns grupos (intervalos) para a maioria das variáveis contínuas (potência, experiência, idade do segurado, idade do veículo) para gerenciar o risco das políticas, e eles foram adotados para a análise. A única exceção é a variável tara (peso). Nesta pesquisa, portanto, os subintervalos foram baseados nas informações da companhia de seguros para todas as variáveis contínuas, exceto para a variável tara. Para ela, percentis (10 a 90) foram empregados para evitar influências subjetivas (consulte também pesquisadores como Laitinen, 1992, e McKee, 2000). A definição ideal de valores limite nos subintervalos geralmente é feita por especialistas de acordo com sua experiência, conhecimentos, hábitos ou convenções (que é a razão pela qual os grupos estabelecidos pela companhia têm sido adotados no artigo) (Dimitras *et al.*, 1999; Slowinski & Zopounidis, 1995). Se não houver um especialista para recodificar as variáveis que poderia seguir sua experiência ou padrões de análise financeira, é desejável evitar a influência subjetiva, na medida do possível (portanto, percentis foram adotados para a variável tara).

Após gravar os fatores de risco, duas tabelas foram obtidas. As duas tabelas de treinamento gravadas, compostas por 4.400 políticas descritas com 13 (sem a variável *bonus-malus*) ou 14 (com a variável *bonus-malus*) fatores de risco e atribuídas a

uma classe de decisão (acidente ou não – 0) foram inseridas em um arquivo de entrada no RSES2. O primeiro resultado obtido pela análise de RS é o cálculo do reduto. Um reduto foi obtido a partir da amostra em ambos os modelos, com ou sem a variável do nível de BM. A única variável que não aparece no reduto é o uso do veículo (PvP) e foi, portanto, eliminada. Consequentemente, embora a teoria de RS seja uma ferramenta muito forte para a seleção de recursos, nesse caso específico a companhia selecionou cuidadosamente as variáveis a fim de se concentrar em poucos fatores de risco para tomar as decisões. Dessa forma, o tempo e o custo do processo de tomada de decisões são minimizados.

Após eliminar a variável redundante nas duas tabelas, o RSES2 induziu dois modelos de regras de decisão (com e sem a variável de BM). Antes de analisar as regras obtidas, ambos os modelos foram validados utilizando duas amostras de teste (1.100 políticas cada um) selecionadas aleatoriamente. Para validar as regras, os dois modelos empregaram a precisão de classificação nas porcentagens dos segurados corretamente classificados. O modelo de RS 1 sem a variável de BM tem uma precisão de classificação média de 72%, ao passo que o modelo de RS 2 com a variável de BM apresenta média de 74,5%.

No geral, os dois modelos são satisfatórios, pois a porcentagem de classificações corretas é superior a 70% e, dessa forma, as regras obtidas para ambos os modelos podem ser interpretadas.

## 5 RESULTADOS E DISCUSSÃO

Primeiro, o modelo de RS 1 sem variável de BM é exibido na Tabela 4. As variáveis são definidas na Tabela 1.

**TABELA 4 – Modelo de RS 1 sem nível de BM**

Regras	TV	USO	PV	TAR	AC	IV	IS	EXP	SEX	NoA	REG	DoG	Classe
1			54-75		Interurbano	5						D	N_A
2						20-30	50-65					G	A
3					Regional			16-20					N_A
4				1010-1078	Interurbano		30-40					G	N_A
5					Interurbano			10-15	F			D	N_A
6			76-118		Urbano		30-40				Múrcia		A
7				1215-1290	Interurbano		30-40						N_A
8					Regional		30-40						N_A
9	CARRO				Urbano		30-40		M		Múrcia		A
10				< 940		20-30	50-65						A
11							30-40		F		Barcelona		N_A
12		Particular				20-30	50-65						A
13	VAN					3							N_A
14								> 20		2		D	N_A
15						3				3			N_A
16			54-75		Interurbano	5			M				N_A
17			54-75			< 1		> 20		5			A
18				9	Interurbano							D	N_A
19						< 1		> 20			Andaluzia	G	A

O modelo de RS 2 com variável de nível de BM é exibido na Tabela 5.

**TABELA 5 – Modelo de RS 2 com nível de BM**

Regras	TV	PV	TAR	AC	IV	IS	EXP	SEX	NoA	REG	DoG	BON	Classe
1				Interurbano								8	N_A
2			< 940		20-30	20						1	A
3				Interurbano	20-30							1	A
4		54-75		Interurbano	5						D		N_A
5				Interurbano							G	4	N_A
6	Van	119-215										1	N_A
7		119-215							3			1	N_A
8					20-30	50-65					G		A
9					20-30	50-65						1	A
10				Regional			16-20						N_A
11				Interurbano				M				6	N_A
12			1010-1078	Interurbano		30-40					G		N_A
13		76-118		Interurbano								4	N_A
14				Interurbano			10-15	F			D		N_A
15		76-118		Urbano		30-40				Múrcia			A
16			1215-1290	Interurbano		30-40							N_A
17				Interurbano							D	6	N_A
18				Regional		30-40							N_A
19	Carro	76-118		Urbano		40-50	> 20	M				2	A
20	Carro			Urbano	30-40			M		Múrcia			A

As regras do primeiro modelo devem ser lidas da seguinte maneira:

- a) Se a potência (PV) = (54-75 HPA) e a área de circulação (AC) = interurbana e a idade do veículo (IV) = (5 anos) e o combustível (DoG) = D (diesel), então N\_A (não acidente).
- b) Se a idade do veículo (IV) = (21 a 30 anos) e a idade do motorista (IM) = (50 a 65 anos) e o combustível (DoG) = G (Gasolina), então A (acidente).
- c) Se a área de circulação (AC) = regional e a experiência (EXP) = (16 a 20 anos), então N\_A (não acidente).
- d) Se... e assim por diante.

Dessa forma, o modelo de RS 2 deve ser lido da mesma maneira.

A comparação das precisões das classificações de ambos os modelos mostra que a variável do nível de BM aumenta a taxa de classificação em 2,5%. Esse resultado está em linha com Bousoño *et al.* (2008). A inclusão de outra variável (BM) implica um aumento de 7,7% no número de variáveis a um aumento de 2,5% na classificação, colocando em questão o valor da inclusão da variável de BM. Tais resultados também estão em conformidade com Hey (1985), uma vez que o BMS não melhora tanto a explicação dos “fatores ocultos”. Na verdade, pesquisas psicológicas mostram variáveis críticas como aversão a riscos, personalidade ou estresse relacionado a incidentes de trânsito.

Existem também estudos empíricos que mostram que homens são duas vezes mais prováveis de serem relatados em, pelo menos, um acidente como motoristas, em comparação às mulheres, e com quase três vezes maior probabilidade de se envolverem em dois ou mais acidentes. Além disso, os motoristas com idade entre 17 e 29 anos têm duas vezes mais probabilidade de relatarem pelo menos um acidente quando comparados àqueles com idade superior a 50 anos (Glendo, Dorn, Davies, Matthews & Taylor, 1996). Considerando esses resultados, dois fatores de risco medidos pelas companhias de seguro são idade e sexo. No entanto, quando os comportamentos de risco foram introduzidos no modelo, o número de

homens ou pessoas entre 17 e 29 anos envolvidos em pelo menos um acidente caiu substancialmente (Turner & McClure, 2003). Por outro lado, a personalidade é uma variável muito importante para prever incidentes automobilísticos (Schwebel, Severson, Ball & Rizzo, 2006). Agressão, tradicionalismo e alienação foram as escalas de personalidade mais frequentemente associadas à direção arriscada e risco de acidente. Acima de tudo, altos níveis de agressão preveem que um motorista pode se envolver em um acidente de carro (Gulliver & Begg, 2007). Finalmente, o estresse do motorista está correlacionado com o envolvimento em acidentes. Além disso, o estresse também está ligado a outras variáveis que poderiam estar correlacionadas a acidentes, como frequência de aborrecimentos diários e agressividade, menos atenção ou estados de humor piores (Matthews, Dorn & Glendon, 1991). Parece improvável que a importância de todas essas variáveis possa ser capturada pelo aumento de 2,5% na precisão da classificação atribuível ao BM no modelo.

Com isso em mente, em razão desses resultados em termos de classificação, as regras de decisão definidas podem ser interpretadas de acordo. Considerando as mais fortes, foram descobertos os seguintes resultados para ambos os modelos:

- a) Nas tabelas 4 e 5, há mais regras para a classe não acidente do que para a classe acidente (Tabela 4 N\_A regras 12- A regras 7; Tabela 5 N\_A regras 13- A regras 7). Algumas regras facilitam a interpretação, pois o modelo é mais compacto e é possível ser mais concreto. Portanto, é mais fácil obter alguns padrões para acidentes de carro em relação à outra classe.
- b) Todas as regras são deterministas. Isso significa que ambas as classes são bem discriminadas entre si. O número de atributos nas regras varia de 2 a 7. Em algumas regras, o modelo de explicação poderia ser definido usando apenas dois fatores.
- c) Os fatores de risco mais relevantes (que apareceram em mais de 50% das regras) para classificar as políticas no modelo de RS 1 são os seguintes: área de circulação,

idade do veículo e idade do motorista. No modelo de RS 2 (com nível de BM), a área de circulação também é o fator de risco mais relevante. O nível de BM é a segunda variável relevante. A variável da área de circulação toma o valor “interurbana” na maior parte das regras que pertencem à classe A (acidente) nos dois modelos. Por outro lado, essa variável ganha o valor “urbana” na maioria das regras referentes à classe N\_A (não acidente). Esse fator mostra que a maioria dos acidentes relatados, que foram provavelmente os mais graves, ocorreram na área de circulação interurbana. Em relação ao nível de BM, ele é atribuído pela companhia com base nas experiências passadas. As regras mostram que os dois melhores níveis (1 e 2) são atribuídos à classe de não acidente. Essa descoberta confirma que a companhia atribui os níveis de bônus corretamente. Isso é especialmente importante dado que o nível de BM é uma variável importante para o cálculo da remuneração.

- d) O fator de risco do uso do veículo não aparece no modelo de RS 2, ao passo que no modelo de RS 1 ele aparece somente em uma regra. Esse fato mostra que o uso do veículo não é um fator determinante, mas pode ser eliminado, pois pertence ao núcleo.
- e) O fator de risco sexo aparece apenas em quatro regras que pertencem às duas classes nos dois modelos. Portanto, não podem ser tiradas conclusões relacionadas a essa variável.

## 6 CONCLUSÕES E PESQUISAS FUTURAS

O objetivo deste trabalho foi testar a validade do uso de níveis “*bonus-malus*” (BM) para classificar satisfatoriamente os segurados. A fim de mostrar essa evidência empírica, um novo método para seguros, a teoria de Rough Set, foi aplicado. De acordo com os dados, a evidência empírica mostra que os fatores de risco comuns

empregados pela companhia de seguros são boas variáveis explicativas para classificar políticas dos segurados.

Além disso, a variável do nível de BM aumenta o poder explicativo dos fatores de risco *a priori*. As diferenças entre os modelos com e sem nível de BM não são muito notáveis. Na verdade, a evidência empírica da amostra demonstra que o poder de explicação do fator de risco com nível de BM é muito baixo (apenas 2,5%). A literatura relacionada consultada mostra, no entanto, que existem muitos fatores importantes que não podem ser considerados *a priori*. Considera-se que essas “características ocultas” são parcialmente reveladas pelo número de reivindicações relatadas pelos segurados, ou seja, o nível de BM. De fato, como mencionado, há muitos fatores relevantes para prever uma personalidade de comportamento perigoso, tomadas de risco ou estresse (entre outros) do motorista que devem ser considerados pelo setor automobilístico.

Para aumentar a capacidade de previsão do nível de BM, questionários psicológicos poderiam ser usados para medir as “características ocultas”. Em termos concretos, os motoristas precisam renovar suas licenças em períodos regulares na Espanha. Isso exige uma avaliação médica que garante as condições físicas para conduzir e que também pode ser usada para testar os valores psicológicos acima mencionados, utilizando, por exemplo, o “Inventário de comportamento do motorista” para estresse ao dirigir (Gulian, Matthews, Glendon, Davies & Debney, 1989); o teste de personalidade “NEO-FFI” (Costa & McCrae, 1992); os “cinco fatores de Zuckerman-Kuhlman” (Zuckerman & Kuhlman, 2000) para a tomada de riscos relacionada à personalidade ou um novo questionário simples que reúna diversos questionários tradicionais. Outra sugestão seria levar em consideração o “sistema de carteira de motorista baseado em pontos”, utilizado em alguns países europeus, como Reino Unido, França, Itália, Irlanda, Luxemburgo e Espanha, como uma procuração para aprimorar a classificação por nível de BM. Há, também, fatores culturais que poderiam afetar o estudo (Nordfaern, Simsekoglu & Rundmo, 2012), como se o motorista presta mais

atenção a informações escritas e sons na estrada ou a informações de tráfego orais e visuais. Outra importante variável é se os motoristas são mais ou menos fatalistas. Portanto, são necessários mais estudos para generalizar os resultados obtidos.

## NOTA

- 1 O software RSES2 foi desenvolvido pelo Instituto de Matemática, Varsóvia, Polônia. Para baixá-lo (Warsaw University, 2005).

## REFERÊNCIAS

- Åberg, L., & Rimmö, P. A. (1998). Dimensions of aberrant driver behavior. *Ergonomics*, 41(1), 39-56.
- Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications*, 18(2), 65-74.
- Arvidsson, S. (2010). Does private information affect the insurance risk? *Working paper, The Geneva Association*, 396, 2010. Retrieved from [http://www.transportportal.se/SWoPEc/Essay\\_1\\_Arvidsson\\_Does\\_private\\_information.pdf](http://www.transportportal.se/SWoPEc/Essay_1_Arvidsson_Does_private_information.pdf)
- Bazan, J., Nguyen, H. S., Nguyen, S. H., Synak, P., & Wróblewski, J. (2000). Rough set algorithms in classification problem. In L. Polkowski, S. Tsumoto, & T. Y. Lin (Eds.), *Rough set methods and applications* (pp. 49-88). New York: Physica-Verlag.
- Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discrimination: An application to corporate failure prediction. *OMEGA: The International Journal of Management Science*, 29(6), 561-576.
- Bousoño, C., Heras, A., & Tolmos, P. (2008). *Factores de riesgo y cálculo de primas mediante técnicas de aprendizaje*. Madrid, España: Ed. MAPFRE.
- Brockett P., Cooper, W., Golden, L., & Pitaktong, U. (1994). A neural network method for obtaining an early warning of insurer insolvency. *The Journal of Risk and Insurance*, 61(3), 402-424.
- Brockett, P., Golden, L., Jang, J., & Yang, C. (2006). A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction. *The Journal of Risk and Insurance*, 73(3), 397-419.
- Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.
- D'Arcy, S. (2005). Predictive modeling in automobile insurance: A preliminary analysis. [Working Paper, 302]. *World Risk and Insurance Economics Congress*, August, Salt Lake City. Retrieved from [http://business.illinois.edu/ormir/Predictive%20Modeling%20in%20Automobile%20Insurance%207-1-05\(PDF\).pdf](http://business.illinois.edu/ormir/Predictive%20Modeling%20in%20Automobile%20Insurance%207-1-05(PDF).pdf)
- Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J. F. (2007). Index, in actuarial modeling of claim counts: Risk classification, credibility and bonus-malus systems. Chichester, UK: John Wiley & Sons.
- Díaz, Z., Segovia, M. J., Fernández, J., & Pozo, E. Machine learning and statistical techniques: An application to the prediction of insolvency in Spanish non-life insurance companies. (2005). *The International Journal of Digital Accounting Research*, 5(9), 1-45. Retrieved from [http://www.uhu.es/ijdar/10.4192/1577-8517-v5\\_1.pdf](http://www.uhu.es/ijdar/10.4192/1577-8517-v5_1.pdf)
- Dimitras, A., Slowinski, R., Susmaga, R., & Zopounidis, C. (1999). Business failure prediction using Rough Sets. *European Journal of Operational Research*, 114(2), 263-280.
- Dionne, G., & Ghali, O. (2005). The bonus-malus system in Tunisia: An empirical Evaluation. *Journal of Risk and Insurance*, 72(4), 609-633.
- Ebanks, B., Karwowski, W., & Ostaszewski, K. (1992). Application of measures of fuzziness to

- risk classification in insurance. *Paper presented at Forth International Conference on Computing and Information ICCI'92*, Toronto.
- Forward, S. (2008). *Driving violations: Investigating forms of irrational rationality*. Uppsala: Universitetsbiblioteket. Retrieved from <http://uu.diva-portal.org/smash/get/diva2:172720/FULLTEXT01>
- Glendon, A. I., Dorn, L., Davies, D. R., Matthews, G., & Taylor, R. G. (1996). Age and gender differences in perceived accident likelihood and driver competences. *Risk Analysis*, 16(6), 755-762. doi: 10.1111/j.1539-6924.1996.tb00826.x
- Goh, C., & Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management*, 24(5), 511-517.
- Greco, S., Matarazzo, B., & Slowinski, R. (1998). A new rough set approach to evaluation of bankruptcy risk. In C. Zopounidis (Ed.), *New operational tools in the management of financial risks* (pp. 121-136). Dordrecht: Kluwer Academic Publishers.
- Greco, S., Matarazzo, B., & Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1), 1-47.
- Gulian, E., Matthews, G., Glendon, A. I., Davies, D. R., & Debney, L. M. (1989). Dimensions of driver stress. *Ergonomics*, 32(6), 585-602.
- Gulliver, P., & Begg, D. (2007). Personality factors as predictors of persistent risky driving behavior and crash involvement among young adults. *Injury Prevention*, 13(6) 376-381.
- Heras, A., Vilar, J. L., & Gil, J. A. (2002). Asymptotic fairness of Bonus-Malus systems and Optimal scales premiums. *The Geneva Papers on risk and Insurance Theory*, 27(1), 61-82.
- Hey, J. (1985). No claim bonus? *The Geneva Papers on risk and Insurance*, 10(36), 209-228.
- Horgby, P.-J. (1998). Risk classification by fuzzy inference. *The Geneva Papers on Risk and Insurance Theory*, 23(1), 63-82.
- International Monetary Fund. (2011). *World economic outlook database*. Retrieved from [www.imf.org](http://www.imf.org)
- Iversen, H. (2004). Risk-taking attitudes and risky driving behavior. *Transportation Research Part F*, 7(3), 135-150.
- Johnson, J. (2006). Can complexity help us better understand risk? *Risk Management*, 8(4), 227-267.
- Kramer, B. (1997). N.E.W.S.: A model for the evaluation of non-life insurance companies. *European Journal of Operational Research*, 98(2), 419-430.
- Laitinen, E. K. (1992). Prediction of failure of a newly founded firm. *Journal of Business Venturing*, 7(4), 323-340.
- Lemaire, J. (1988). A comparative analysis of most European and Japanese Bonus-malus Systems. *Journal of Risk and Insurance*, 55(4), 660-681.
- Lemaire, J. (1990). Fuzzy insurance. *ASTIN Bulletin*, 20(1), 33-55.
- Lemaire, J. (1995). *Bonus-malus systems in automobile insurance*. Boston: Kluwert Academic Publisher.
- Martinez de Lejarza Esparducer, I. (1996, September). Forecasting company failure: Neural approach versus discriminant analysis: An application to Spanish insurance companies of the 80's. *International Conference on Artificial Intelligence in Accounting, Finance and Tax*, Punta Umbria (Huelva), Spain, 2.
- Matthews, G., Dorn, L., & Glendon, A. (1991). Personality correlates of driver stress. *Personality and Individual Differences*, 12(6), 535-549.
- McKee, T. (2000). Developing a bankruptcy prediction model via rough sets theory.



*International Journal of Intelligent Systems in Accounting, Finance and Management*, 14(3), 159-173.

Nordfjaern, T., Simsekoglu, O., & Rundmo, T. (2012). A comparison of road traffic culture, risk assessment and speeding predictors between Norway and Turkey. *Risk Management*, 14(3), 202-221.

Nurmi, H., Kacprzyk, J., & Fedrizzi, M. (1996). Probabilistic, fuzzy and rough concepts in social choice. *European Journal of Operational Research*, 95(2), 264-277.

Park, S., Lemaire, J., & Chua, C.T. (2009). Is the design of Bonus-Malus Systems influenced by insurance maturity or national culture? Evidence from Asia. *The Geneva Papers*, 35(S1), 7-27.

Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Dordrecht: Kluwer Academic Publishers.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R., & Ziarko, W. (1995). Rough Sets. *Communications of the ACM*, 38(11), 89-97. Retrieved from [http://dl.acm.org/ft\\_gateway.cfm?id=277421&ftid=17537&dwn=1&CFID=220789019&CFTOKEN=72446287](http://dl.acm.org/ft_gateway.cfm?id=277421&ftid=17537&dwn=1&CFID=220789019&CFTOKEN=72446287)

Pitrebois, S., Denuit, M., & Walhin, J.F. (2006). Multi-event Bonus-malus scales. *The Journal of Risk and Insurance*, 73(3) 517-528.

PwC. (2012). *The five keys to the industry*. Retrieved from <http://www.pwc.es/en/financiero-seguros/claves-sector-seguros.jhtml>

Resende, P. C., Jr., & Guimãres, T. (2012). Inovação em serviços: O estado da arte e uma proposta de agenda de pesquisa. *Revista Brasileira de Gestão de Negócios*, 14(44), 293-313.

Richadeau, D. (1999). Automobile insurance contracts and risk of accident: An empirical test using French individual data. *The Geneva Papers on Risk and Insurance Theory*, 24(1), 97-114.

Salcedo Sanz, S., Fernández Villacañas, J. L., Segovia Vargas M. J., & Bousoño Calzón, C. (2005). Genetic programming for the prediction of insolvency in non-life insurance companies. *Computers and Operations Research*, 32(4), 749-765.

Salcedo Sanz, S., Prado Cumplido, M., Segovia Vargas, M. J., Perez Cruz, F., & Bousoño Calzón, C. (2004). Feature selection methods involving Support Vector Machines for prediction of insolvency in non-life insurance companies. *Intelligent Systems in Accounting, Finance and Management*, 12(4), 261-281.

Sanchís, A., Segovia, M. J., Gil, J. A., Heras, A., & Vilar, J. L. (2007). Rough Sets and the Role of Monetary Policy in Financial Stability (Macroeconomic Problem) and the Prediction of Insolvency in the Insurance Sector (Microeconomic Problem). *European Journal of Operational Research*, 181(3), 1554-1573.

Schwebel, D. C., Severson, J., Ball, K.K., & Rizzo, M. (2006). Individual difference factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation-seeking. *Accident Analysis and Prevention*, 38(4), 801-810.

Segovia-González, M. M., Contreras, I., & Mar-Molinero, C. A. (2009). DEA analysis of risk, cost, and revenues in insurance. *Journal of Operational Research Society*, 60(11), 1483-1494.

Segovia-Vargas M. J., Salcedo-Sanz, S., & Bousoño-Calzón, C. (2004). Prediction of Insolvency in non-life insurance companies using support vector machines and genetic algorithms. *Fuzzy Economic Review*, 9(1), 79-94.

Slowinski, R., & Zopounidis, C. (1995). Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 4(1), 27-41.

Shapiro, A. (2005). Fuzzy logic in insurance: the first 20 years. *Actuarial Research Clearing*

*House*, 39(1), 1-32. Retrieved from <https://www.soa.org/News-and-Publications/Publications/Proceedings/Arch/pub-arch-table-of-contents-2005-1.aspx>

Shen, Q., & Jensen, R. (2007). Rough sets, their extensions and applications. *International Journal of Automation and Computing*, 4(3), 217-228.

Shyng, J.-Y., Wang, F.-K., Tzeng, G.-H., & Wu, K.-S. (2007). Rough Set Theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications*, 32(1), 56-64.

Silva, J. C. B. (2004). A escolha da seguradora para o seguro fiança locatícia na óptica dos corretores de seguros. *Revista Brasileira de Gestão de Negócios*, 6(15), 49-68.

Skowron, A., & Rauszer, C. M. (1992). The discernibility matrices and functions in information systems. In R. W. Slowinski (Ed.), *Intelligent decision support* (Chap. 2, pp. 331-362). Dordrecht: Kluwer Academic Publishers.

Spanish National Institute of Statistics (2015). *Accidentes. Serie 2004-2012*. Retrieved from <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t10/a109/a04/&file=pcaxis>

Turner, C., & McClure, R. (2003). Age and gender differences in risk-taking behavior as an explanation for high incidence of motor vehicle crashes as a driver in young males. *Injury Control and Safety Promotion*, 10(3), 123-130.

Warsaw University. (2005). *RSES 2.2 User's Guide*. Retrieved from [http://logic.mimuw.edu.pl/~rses/RSES\\_doc\\_eng.pdf](http://logic.mimuw.edu.pl/~rses/RSES_doc_eng.pdf)

Wit, G. W. (1982). Underwriting and uncertainty. *Insurance: Mathematics and Economics*, 1(4), 277-285.

Witlox, F., & Tindemans, H. (2004). The application of rough sets analysis in activity-based modeling, opportunities and constraints. *Expert Systems with Application*, 27(4), 585-592.

Young, V. (1996). Insurance rate changing: A fuzzy logic approach. *Journal of Risk and Insurance*, 63(3), 461-484.

Zuckerman, M., & Kuhlman, M. (2000). Personality and Risk-Taking: Common bisocial factors. *Journal of Personality*, 68(6), 999-1029.