



Programa de Certificação

Categoria: EYU Certificate Advanced Level

Orientador: Anderson Silva

GESTÃO DE FRAUDE NO MERCADO DE SEGUROS DE AUTOMÓVEL – GANHO DE PERFORMANCE E COMPETITIVIDADE PARA SEGURADORAS

Profissional: Enrico Davite Neto
Rank: Senior
SL: Advisory Services
Escritório: Belo Horizonte
SMU: Risk / Actuarial Services

Palavras-Chave:

Coefficiente de regressão: (B_x): Valor numérico da estimativa do parâmetro diretamente associado com uma variável independente, por exemplo, $Y = B_0 + B_1X_1$, o valor B_1 é o coeficiente da regressão para a variável X_1 . O coeficiente de regressão representa o montante de variação na variável dependente em relação a uma unidade de variação na variável independente.

Escore: Escore ou coeficiente discriminante tem seu peso determinado pela estrutura de variância das variáveis originais nos grupos de variáveis dependentes. Variáveis independentes com grande poder discriminatório geralmente têm grande peso, e as que apresentam pouco poder discriminatório geralmente têm pesos pequenos; no entanto, a multicolinearidade entre as variáveis independentes provoca exceção a regra.

Escore de Corte: Critério (escore) contra o qual cada escore Z discriminante individual é comparado para determinar a pertinência prevista em um grupo. Quando a análise envolve dois grupos, a previsão de grupo é determinada computando-se um único escore de corte. Neste artigo o escore de corte foi determinado como 0,5.

Escore Z Discriminante: Escore definido pela função discriminante para cada objeto na análise e geralmente dado em termos padronizados. Também conhecido como escore Z , é calculado para cada objeto em cada função discriminante e usado em conjunção com o escore de corte para determinar pertinência prevista ao grupo.

Erro Tipo I: Probabilidade de rejeitar uma hipótese nula quando ela deveria ser aceita, ou seja, concluir que duas médias são significativamente diferentes quando de fato são as mesmas.

Intercepto: (B_0): Valor do eixo Y (eixo da variável independente) onde a reta definida pela equação $Y = B_0 + B_1X_1$ cruza o eixo. É descrito pelo termo constante B_0 na equação de regressão. Além de seu papel na previsão, o intercepto pode ter uma interpretação gerencial. Se a completa anuência da variável independente tem significado, então o intercepto representa essa quantia.

Multicolinearidade: Grau em que uma variável pode ser explicada pelas outras variáveis na análise. Quando a multicolinearidade aumenta, ela complica a interpretação da variável estatística por ser mais difícil de determinar o efeito de qualquer variável individual, devido às inter-relações entre elas.

Nível de Significância Nominal: Nível de significância associado ao teste estatístico das diferenças entre dois ou mais grupos. Normalmente, valores pequenos, como 0,05 ou 0,01, são especificados para minimizar a possibilidade de se cometer um erro tipo I.

Odds: Razão de chances

Probabilidade de Significância (P-Valor): É a probabilidade de se ter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra, sob a hipótese nula. Por exemplo, em testes de hipótese, pode-se rejeitar a hipótese nula a 5% caso a probabilidade de significância seja menor que 5%.

Regressão Logística: Forma especial de regressão na qual a variável dependente é não-métrica, dicotômica, binária. Apesar de algumas diferenças, a maneira geral de interpretação é semelhante à da regressão linear.

Sumário

1. Introdução	5
2. Problema de Pesquisa	5
3. O Mercado de Seguros de Automóvel no Brasil	7
3.1 Produção de Prêmio Automóvel vs Demais Ramos	8
3.2 Sinistralidade.....	9
3.3 Resultado Operacional do Mercado Segurador	9
4. Fraude no Mercado de Seguros de Automóvel	10
4.1 Números recentes sobre a prática da Fraude.....	11
5. Aplicação da Análise Discriminante como recurso para detecção de perfis de segurados fraudulentos	12
5.1 Regressão Logística.....	13
5.2 Análise da Qualidade de Ajuste do Modelo.....	14
6. Coleta, tratamento e análise estatística dos dados	15
7. Resultado da Pesquisa	20
7.1 Validação do Resultado.....	26
8. Conclusão	32
9. Referências	33

1. Introdução

O contrato de seguro é norteado pela boa-fé, que pode ser descrita como a base ética da ordem jurídica e do convívio em sociedade, em prol da viabilidade dos negócios, correta precificação e saúde financeiras das provisões técnicas (CNseg, 2012).

A boa-fé é absolutamente antagônica à fraude contra o seguro, uma vez que além de viciar o negócio, também se presta ao estímulo da criminalidade e redução da propensão ao consumo de seguro em prejuízo desse importante mercado e de todos os consumidores de seguros.

Além de perdas morais, a fraude traz perdas financeiras e pessoais que muitas vezes não podem ser recuperadas. É necessário que os segurados saibam que são eles quem pagam essas contas geradas pela fraude. O seguro pode custar muito menos se houver uma consciência maior das pessoas quanto ao uso correto do produto (Avellar, 2003).

A fraude ocorre no momento da celebração do contrato de seguro, durante a sua vigência ou, sobretudo, perante a ocorrência dos sinistros (Brites, 2006). Nos sinistros, a fraude é caracterizada quando adulterações nas informações e documentações passadas à seguradora são feitas com intenção de proveito por parte do segurado.

Combater a fraude é uma atitude ética e inerente à própria atividade seguradora. Acreditando nisso, o mercado brasileiro engaja-se, desde 2004, num programa institucional de quantificação da fraude, que vem sendo realizado através do SQF - Sistema de Quantificação da Fraude, trilhando o caminho da educação, sensibilização dos Órgãos Públicos e disseminação da cultura do seguro (CNseg, 2012).

Com base neste contexto, este artigo propõe um estudo das variáveis que consolidam a composição dos sinistros fraudulentos, para então propor um modelo discriminante com o objetivo de identificar os sinistros com indícios de fraude.

2. Problema de Pesquisa

O contrato de seguros é, em tese, um contrato de boa-fé entre duas partes que tem como objetivo dar cobertura a um determinado bem, que pode ser uma vida ou um objeto, contra

um infortúnio, visto que qualquer bem possui um valor monetário, e que no caso de vida, explica-se pela perda de renda para a família (ALVIM, 1999).

As operações de seguros, aqui entendidas como uma transferência de responsabilidades que as Companhias Seguradoras realizam no seu dia a dia, obedece a critérios estabelecidos pelo seu órgão regulador, a SUSEP – Superintendência de Seguros Privados (XAVIER, 1999). Como especialistas, cabe aos atuários, à missão de estabelecer uma equação de equilíbrio que viabilize essa operação. Em síntese, o processo de construção do preço de um produto, sob o ponto de vista do risco, consiste na avaliação desse risco, passando pelas características do produto que se pretende oferecer e, finalmente, pela capacidade de prever os riscos futuros. Por outro lado, torna-se necessário estabelecer uma ponderação sobre diversos outros aspectos inerentes ao produto, tais como: público alvo a que se destina, nível de renda da população que irá adquirir o produto e o nível cultural e de necessidade (Ernst & Young, 2003). Para compor o processo de definição de preços, baseamo-nos nos acontecimentos passados, estatísticas de mercado e principalmente, na experiência própria da seguradora. Se esses acontecimentos, por sua vez, estiverem em desacordo com uma das características básicas do risco segurável, que neste contexto seria independência da vontade das partes, ou seja, sem fraude, causarão distorções nos preços do seguro (BRITES, 2006).

A fraude significa engano lesivo, mentira, abuso de confiança, ação de má-fé, adulteração, ou seja, é crime e deve ser vista como ato criminoso contra a economia, pois distorce a relação de equilíbrio das operações de seguros, elevando artificialmente a sinistralidade¹, ou seja, afeta diretamente o preço do seguro (Ernst & Young, 2003). Uma visão realista do combate à fraude, conforme afirmado pelo professor Marcelo Lessa Bastos em palestra ministrada em 2005 a Fenaseg é:

“O combate à fraude é, antes de tudo, uma medida de proteção ao consumidor honesto, eis que o custo da fraude é ‘transferido’ para o consumidor. Aliás, as seguradoras, com isto, demonstram que aplicam muito bem a ‘doutrina do seguro’, exposta pelo professor Peon de Sá², já que transferem um risco que deveria ser seu – a fraude – para o consumidor, através da elevação dos prêmios. (Bastos, 2004)

Com base nessas condições, percebe-se que o papel da seguradora consiste em repassar aos segurados o acréscimo gerado nos preços de seguros em função das fraudes.

¹ Sinistralidade é, segundo Xavier, a razão (divisão) do prêmio sobre o sinistro que gera uma margem utilizada para precificar o seguro.

² Peon de Sá - Não há e nunca houve “interpretação inadequada de cláusulas contratuais” firmadas pelos segurados por parte do Ministério Público e do Poder Judiciário. O que há às vezes é, adequação dessas cláusulas às Leis Maiores – Constituição Federal e Código de Defesa do Consumidor – que impõem limites ao pacta sunt servanda. Cabe ao Juiz, se for o caso, assegurar o pleno gozo dos direitos individuais, a um só indivíduo, ainda que isto implique em prejudicar toda a coletividade.

Neste contexto, questiona-se como problema central da pesquisa: como um modelo discriminante pode auxiliar na identificação dos processos fraudulentos nas áreas operacionais das seguradoras, melhorando assim, a situação dos bons consumidores de seguros?

3. O Mercado de Seguros de Automóvel no Brasil

O mais popular de todos os seguros, o seguro de automóveis tornou-se um item básico, a ponto de ser incluído entre os custos totais quando se pensa na compra de um carro. Isso se dá principalmente nas grandes capitais, onde a probabilidade de roubo, furto ou acidentes envolvendo automóveis é maior. O índice de sinistralidade, que resulta na divisão dos valores pagos em indenizações pelos arrecadados em prêmios, da carteira de automóveis das Companhias Seguradoras brasileiras é pouco mais de 65,19% (SUSEP, 2012). Isso significa que, de cada R\$100,00 que os segurados pagam em prêmio as Companhias Seguradoras gastam R\$65,19 com indenizações. A elevada sinistralidade faz com que o seguro de automóveis seja, proporcionalmente, um dos mais caros. Como o risco de sinistralidade é alto, as seguradoras apertam nos preços e nas franquias. E o resultado é que apenas aproximadamente 22,5% da frota brasileira está segurada (Consultoria Siscorp, 2011).

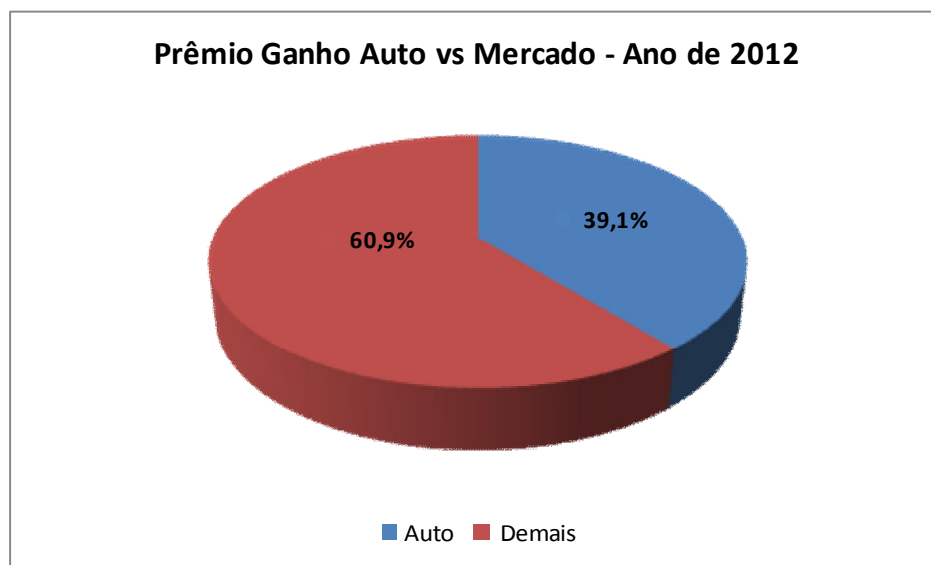
Atualmente, segundo SUSEP, existem 116 seguradoras em operação no mercado de seguros brasileiro. Dentre elas, apresentamos a seguir tabela com as 10 maiores considerando o volume de prêmios ganhos no ano de 2012 para o grupo de automóvel.

Empresa	Prêmio Ganho (R\$)
PORTO SEGURO COMPANHIA DE SEGUROS GERAIS	3.384.889.934
BRANCO AUTO/RE COMPANHIA DE SEGUROS	3.169.524.033
MAPFRE SEGUROS GERAIS S.A.	2.968.316.123
SUL AMÉRICA COMPANHIA NACIONAL DE SEGUROS	2.282.573.442
ITAÚ SEGUROS DE AUTO E RESIDÊNCIA S.A.	1.664.468.368
HDI SEGUROS S.A.	1.569.413.883
ALLIANZ SEGUROS S.A.	1.443.318.163
LIBERTY SEGUROS S.A.	1.342.074.463
AZUL COMPANHIA DE SEGUROS GERAIS	1.181.726.415
TOKIO MARINE SEGURADORA S.A.	743.297.829

Fonte: SUSEP 2012

3.1 Produção de Prêmio Automóvel vs Demais Ramos

A representatividade da produção de prêmios do seguro de automóvel frente aos demais ramos no ano de 2012 pode ser vista no gráfico a seguir. Contemplam no agrupamento de demais ramos os seguintes grupos de ramos: Patrimonial, Riscos Especiais, Responsabilidades, Transportes, Riscos Financeiros, Pessoas Coletivo, Habitacional, Rural, Outros, Pessoas individual, Marítimos e Aeronáuticos.



Fonte: SUSEP 2012

Conforme se observa, 39,1% de toda produção do mercado no ano de 2012 refere-se a automóvel. Historicamente essa representatividade apresentou o seguinte cenário:

	2011	2010	2009	2008
Auto	43%	48%	47%	47%
Demais	57%	52%	53%	53%

Fonte: SUSEP

De acordo com a tabela acima o volume de prêmios ganhos manteve-se estável entre os anos de 2008 e 2010 apresentando leve redução entre 2011 e 2012. Em uma análise univariada da representatividade percentual, poderíamos dizer que o mercado de automóvel apresentou uma leve redução na produção de prêmios frente aos demais ramos. No entanto, se avaliarmos a tabela a seguir que demonstra o volume de prêmios ganhos ao longo dos cinco anos, percebemos que, com exceção aos anos de 2011 para 2012, auto e demais ramos apresentaram crescimento. Logo podemos concluir que o mercado de seguros vem apresentando crescimento em todos os ramos de seguros.

	2012	2011	2010	2009	2008
Auto	22.735.017.321	23.900.129.285	21.482.545.992	18.656.833.386	16.527.906.983
Demais	35.395.087.595	31.596.265.985	23.657.946.130	21.373.442.732	18.553.338.557

Fonte: SUSEP

3.2 Sinistralidade

O comportamento da sinistralidade do grupo de automóvel ao longo dos cinco anos avaliados pode ser visto na tabela a seguir.

	2012	2011	2010	2009	2008
Auto	65%	69%	67%	69%	68%

Fonte: SUSEP

Conforme se pode observar, a sinistralidade manteve-se entre o intervalo de 65% e 69% nos cinco anos avaliados. Se observarmos a tabela a seguir percebemos o quanto o grupo de automóvel tem a sinistralidade mais elevada quando comparada aos demais ramos.

	2012	2011	2010	2009	2008
Demais	30%	29%	35%	38%	40%

Fonte: SUSEP

3.3 Resultado Operacional do Mercado Segurador

O resultado operacional é valor obtido pelo confronto de todas as receitas, custos e despesas que têm relação direta com a existência da companhia seguradora, ou seja, todos os eventos relacionados à atividade fim. Apura-se o resultado operacional deduzindo-se do resultado bruto as despesas do período e o resultado das transações financeiras. Seu resultado possibilita medir a eficiência operacional da companhia seguradora.

O resultado operacional obtido pelo mercado segurador entre os anos de 2008 a 2012 é apresentado na tabela a seguir:

Período	Resultado Operacional
2012	18.745.469.181
2011	17.013.751.367
2010	14.847.771.090
2009	11.860.551.870
2008	11.209.324.628

Fonte: SUSEP

Conforme se observa, o crescimento foi verificado ao longo dos últimos quatro anos. Essa realidade se mostra positiva em vários aspectos para o mercado uma vez que no caso de automóvel, menos de 50% da frota brasileira está segurada (Consultoria Siscorp, 2011).

4. Fraude no Mercado de Seguros de Automóvel

A indústria seguradora, segundo Brites (2006), depara-se frequentemente com tentativas de fraude que ocorrem no momento da celebração do contrato de seguro, durante a sua vigência ou, sobretudo, perante a ocorrência dos sinistros. Infelizmente, boa parte dessas tentativas acaba por concretizar-se na ocorrência dos sinistros.

As fraudes concretizadas, ou seja, aquelas que a seguradora não consegue provar a má fé do segurado, acabam invariavelmente gerando um prejuízo patrimonial para as mesmas que, por consequência, levam os valores desses prejuízos no cálculo dos prêmios de seguros, o que traduz, na prática, em custos adicionais para toda a massa de segurados.

A fraude é vulgarmente definida como toda ação ou omissão ilegítima levada a cabo por qualquer um dos intervenientes num sinistro, com o objetivo de conseguir um benefício próprio ou de favorecer um terceiro (Brites, 2006).

Os tipos de fraude mais comuns na ocorrência do sinistro são:

- Simulação de sinistros (furto, roubo, incêndio) e participações a favor;
- alteração das circunstância do sinistro de forma que a responsabilidade do evento seja de conveniência do envolvido;
- falsa documentação de suporte às despesas suportadas com o objetivo de elevar as indenizações a receber;
- danos provocados intencionalmente com o objetivo de agravá-los;
- duplicação de reclamações com algum intervalo de tempo ou em diferentes seguradoras.

As fraudes cometidas e concretizadas têm consequências diretas que se traduzem genericamente em: prejuízos patrimoniais para as seguradoras, custos adicionais para os segurados num acréscimo de prêmio, perda de competitividade. Torna-se necessário combater a calamidade, cometida por uns poucos e suportada por todos, pois os segurados que agem de boa fé têm o direito de ser protegido pela seguradora contra possíveis fraudadores e injustiçadas pretensões. Não se pode recobrar a comunidade de segurados as quantias injustificadas que são pagas por sinistros fraudulentos. As entidades seguradoras que não investigam as possíveis fraudes, tramitando constantemente de forma generosa os seus sinistros, constituem uma verdadeira tentação para os fraudadores. É necessário a criação de mecanismo para detectá-las. Este artigo propõe um possível mecanismo a fim de atingir este objetivo.

4.1 Números recentes sobre a prática da Fraude

Por entender que combater a fraude é uma atitude ética e inerente à própria atividade seguradora, o mercado brasileiro através da CNseg, engaja-se desde 2004 num programa institucional de quantificação da fraude.

No último relatório com estatísticas de 2011 divulgado ao mercado, os sinistros com suspeitas de fraude somaram R\$1,86 bilhões o que representou 7,2% do valor total dos sinistros do universo pesquisado (R\$ 25,9 bilhões). As fraudes comprovadas somaram cerca de R\$ 338 milhões o que representou 1,3% do valor total de sinistros e 0,8% do prêmio ganho (CNseg, 2012).

Em uma comparação com os números divulgados no relatório de 2010, percebeu-se um aumento de 25% no total dos sinistros avisados e de 12% nos sinistros investigados. No comparativo também foi verificado que apesar do aumento no valor total dos sinistros avisados, o valor total dos sinistros suspeitos permaneceu constante, no entanto houve um aumento de 18% no valor total de sinistros com fraude comprovada (CNseg, 2012).

Todo esse montante pago em indenizações com sinistros fraudulentos impacta diretamente o preço do seguro de automóvel que será praticado no mercado em períodos posteriores. A definição do preço do seguro leva em consideração além a experiência que as seguradoras tiveram com pagamento de sinistros, outras variáveis como despesas administrativas, comerciais e lucro esperado. Como os sinistros fraudulentos impactam com sua representatividade no pagamento de indenizações e despesas administrativas, o mesmo reflete de maneira a elevar o preço do seguro a ser praticado em períodos posteriores. Logo, os bons consumidores acabam por “financiar” o elevado custo que a fraude gera para todo o mercado.

5. Aplicação da Análise Discriminante como recurso para detecção de perfis de segurados fraudulentos

Em qualquer decisão que tomamos em nossas vidas, sempre levamos em conta um grande número de fatores. Obviamente nem todos estes pesam da mesma maneira na hora de uma escolha. Às vezes, por tomarmos uma decisão usando a intuição, não identificamos de maneira sistemática estes fatores, ou seja, não identificamos quais as variáveis que afetaram a nossa decisão.

Segundo Hair et. al., (2005), estabelecer relações, encontrar ou propor leis explicativas é o papel próprio da ciência. Para isso é necessário controlar, manipular, medir as variáveis que são consideradas relevantes ao entendimento do fenômeno analisado.

Todos nós somos treinados a analisar as variáveis isoladamente e, a partir desta análise, fazer inferências sobre a realidade (SILVEIRA, 2003). Esta simplificação tem vantagens e desvantagens. Quando um fenômeno depende de muitas variáveis, geralmente este tipo de análise falha, pois não basta conhecer informações estatísticas isoladas, mas é necessário também conhecer a totalidade destas informações fornecida pelo conjunto das variáveis (HAIR et. al., 2005).

O desenvolvimento tecnológico proveniente das descobertas científicas tem alavancado o próprio desenvolvimento científico, ampliando em larga escala a capacidade de se obter informações de acontecimentos e fenômenos que estão sendo analisados (HAIR, et. al., 2005). Uma grande massa de informação deve ser processada antes de ser transformada em conhecimento. Portanto, cada vez mais estamos necessitando de ferramentas estatísticas que apresentem uma visão mais global do fenômeno, que aquela apresentada numa abordagem univariada. Segundo Hair et. al. (2005) a denominação Análise Multivariada, corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos.

Segundo Mingoti (2005) a análise estatística multivariada se divide em dois grupos, sendo o primeiro baseado em técnicas exploratórias de sintetização da estrutura de variabilidade dos dados e o segundo consistindo em técnicas de inferência estatística. Cada grupo apresenta uma série de métodos com suas respectivas características. Por exemplo, para o primeiro grupo a grande maioria dos métodos independem do conhecimento da forma da distribuição de probabilidade das variáveis medidas nos elementos amostrais. Já no segundo grupo encontram-se métodos de estimação de parâmetros, testes de hipóteses, análise de variância, de covariância, dentre outros. Os métodos que compõe o primeiro grupo são: a análise de componentes principais, análise fatorial, análise de correlações canônicas, análise de agrupamentos, análise discriminante e análise de correspondência. Como no estudo

desenvolvido neste artigo utilizaremos a técnica de análise discriminante, nos enquadraremos no primeiro grupo.

5.1 Regressão Logística

A regressão logística é uma técnica que visa descrever a natureza do relacionamento entre um determinado evento, que neste estudo é a ocorrência da fraude, e a (as) variáveis independentes ou regressoras.

Muitas funções têm sido propostas para se explicar o comportamento de uma variável resposta dicotômica e a regressão logística é amplamente usada por duas razões:

- do ponto de vista matemático, é uma função facilmente manipulável, já que utiliza uma transformação logística;
- os coeficientes da regressão logística têm um significado de simples interpretação.

Para simplificar a notação, neste artigo usaremos $\Pi(Z|X)$ para representar a probabilidade de que o sinistro seja fraudulento ($Z=1$) dado os valores das variáveis regressoras $X = [X_1 X_2 \dots X_p]$. A estimação da probabilidade de um determinado sinistro pertencer à população dos fraudulentos é dada pela seguinte formulação:

$$\Pi(1|X) = \frac{\exp(\beta_0 + \beta'x)}{1 + \exp(\beta_0 + \beta'x)}$$

Onde:

β_0 = Constante, que é chamada de intercepto,

$\beta'x$ = estimativa do parâmetro diretamente associado com a variável independente, que é chamada de coeficiente da regressão

Já para os sinistros da população dos não fraudulentos ($Z=0$), a estimativa é dada pela formulação a seguir:

$$\Pi(0|X) = \frac{1}{1 + \exp(\beta_0 + \beta'x)}$$

Onde:

β_0 = Constante, que é chamada de intercepto,

$\beta'x$ = estimativa do parâmetro diretamente associado com a variável independente, que é chamada de coeficiente da regressão

sendo $\Pi(1|X) + \Pi(0|X) = 1$ e os parâmetros do modelo, $\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$, estimados pelo método estatístico de máxima verossimilhança (Mingoti, 2005; Hair et. al, 2005). Após o ajuste do modelo logístico, os escores de cada elemento amostral das populações dos sinistros fraudulentos ($Z=1$) e não fraudulentos ($Z=0$) são calculados, sendo cada elemento classificado como pertencente à população cuja probabilidade $\Pi(Z)$ é maior, $Z = 1,0$.

5.2 Análise da Qualidade de Ajuste do Modelo

A análise da qualidade do ajuste do modelo pode ser realizada pelos seguintes métodos:

1) Método da Resubstituição: Neste método, os escores de cada elemento amostral observado na população dos sinistros fraudulentos e não fraudulentos são calculados, sendo a regra de discriminação utilizada para classificar os $n = n_1 + n_2$ elementos da amostra conjunta. Quando a função discriminante é de boa qualidade, espera-se que ela apresente uma grande porcentagem de acerto na classificação dos elementos amostrais em relação à população a que de fato pertencem (MINGOTI, 2005).

2) Método de colocação de elementos à parte para classificação: Neste método, a amostra conjunta de $n = n_1 + n_2$ elementos é repartida em duas partes, uma que vai servir para a construção da regra de discriminação (amostra de treinamento) e outra que vai ser utilizada para a estimação das probabilidades de classificações incorretas (amostra de validação). Inicialmente, seleciona-se aleatoriamente, alguns sinistros das amostras das populações dos sinistros fraudulentos e dos não fraudulentos, deixando-os a parte da amostra original de $n_1 + n_2$ elementos. Para cada um destes elementos, sabe-se qual a população que ele pertence e, portanto, eles servirão para testar a função discriminante construída a partir dos elementos amostrais restantes. A regra de discriminação estimada é utilizada para classificar os elementos que foram colocados à parte inicialmente, e as proporções de classificações incorretas são calculadas da mesma forma como a descrita no método da resubstituição. A desvantagem do método é a redução do tamanho da amostra original para a estimação da regra de discriminação, o que poderá diminuir acentuadamente a confiabilidade da regra de classificação construída, se as amostras não forem grandes (MINGOTI, 2005).

3) Método de Lachenbruch (Validação Cruzada): Este procedimento é muito utilizado e está implementado em praticamente todos os softwares estatísticos. A validação cruzada consiste nos seguintes passos:

1º passo: retira-se um vetor de observações da amostra conjunta e utilizam-se os $(n_1 + n_2 - 1)$ elementos amostrais restantes para construir a função de discriminação;

2º passo: utiliza-se a regra de discriminação construída no 1º passo para classificar o elemento que ficou à parte da construção da regra de discriminação, verificando se a regra de discriminação conseguiu acertar na sua real procedência ou não.

3º passo: retorna-se o elemento amostral que foi retirado no 1º passo à amostra original e retira-se um outro elemento amostral diferente do primeiro. O 1º e 2º passos são repetidos.

Os passos 1,2,3 devem ser repetidos para todos os $(n_1 + n_2)$ elementos da amostra conjunta e as probabilidades dos erros do tipo 1 e 2 de classificações incorretas serão estimadas respectivamente por:

$$\hat{p}(2|1) = \frac{n_{12}}{n_1} \quad \text{e} \quad \hat{p}(1|2) = \frac{n_{21}}{n_2}$$

sendo n_{12} e n_{21} como definidos anteriormente. As estimativas deste método são aproximadamente não viciadas e melhores que o método da ressubstituição para populações normais e não normais (MINGOTI, 2005).

Neste artigo utilizaremos o método de colocação de elementos à parte para classificação para avaliarmos a qualidade de discriminação do modelo ajustado.

Os resultados do ajuste do modelo logístico são apresentados a partir da seção 6. As variáveis estudadas foram modelo do veículo, valor de mercado do veículo, ano do veículo, franquia, região, número de prestações, corretor, estado civil, garagem, sexo, idade, vigência e tipo de perda. A descrição de cada uma delas é apresentada no capítulo III na parte de tratamento dos dados.

6. Coleta, tratamento e análise estatística dos dados

Os dados necessários para realização deste artigo foram coletados de uma seguradora de médio porte do mercado mineiro. Extraímos 186 registros de sinistro do ramo de automóvel dos quais classificamos como fraudulentos e 600 registros de sinistros que classificamos como não fraudulentos. Por sinistros fraudulentos definimos que foram os casos que em seu processo de regulação sofreram sindicância e ao ser encerrado, não houveram registro de indenização paga ao segurado. Por sinistro não fraudulentos definimos que foram aqueles casos que em seu processo de regulação sofreram sindicância e ao ser encerrado,

apresentaram registros de indenização paga ao segurado. Os períodos de análise dos sinistros são referentes aos anos de 2004, 2005 e 2006.

Após o término da coleta de dados, foi realizada uma análise estatística dos mesmos para que se tornasse possível organizá-los e compreendê-los, através de um processo em que se procurou dimensões, padrões e relações. É um processo complexo, pois implica organização e interpretação.

Para identificação dos sinistros que caracterizamos como fraudulentos, utilizamos os registros de controle que a área operacional da seguradora nos disponibilizou. Neste momento obtivemos uma quantidade pequena de registros, pois a seguradora não apresentava métodos eficazes de segurança contra a fraude, ou seja, não havia muitos casos onde se conseguia provar a “má fé” do segurado, mesmo à seguradora sabendo das intenções do mesmo. Para determinar se o sinistro é ou não fraudulento, a seguradora conta com análise de técnicos e avaliação pericial de empresas terceirizadas. Quando o técnico desconfia da “má fé” do segurado, é acionado o serviço técnico de sindicância, onde neste momento, é realizado um estudo do caso. No fim do estudo é emitido um relatório que dará ao técnico subsídios para provar a intenção do segurado. Na prática, existem muitos casos onde o técnico tem o relatório do perito indicando a intenção da fraude, mas por muitas vezes não ser este um documento suficiente para provar a intenção do segurado em juízo, a seguradora acaba pagando a indenização. Este artigo tem por objetivo oferecer mais um recurso visando minimizar a ação dos segurados de “má fé”, uma vez que ela estabelece a discriminação do segurado no momento da contratação do seguro.

Para identificação dos sinistros que caracterizamos como não fraudulentos, utilizamos os registros do departamento estatístico da seguradora. Neste momento obtivemos uma quantidade imensa de informações das quais selecionamos uma amostra de 600 observações seguindo os seguintes critérios:

- Região de circulação: selecionamos os sinistros que tinham como região de circulação os estados de Minas Gerais, Rio de Janeiro e São Paulo;
- Valor do sinistro: selecionamos os casos onde o valor do sinistro era superior ao da franquia do segurado.

Na base de dados coletada, estudamos treze variáveis das quais receberam a classificação de: modelo do veículo, valor de mercado do veículo, ano do veículo, tipo de franquia, região de circulação, número de prestações, código do corretor, estado civil, garagem comercial, sexo, idade, vigência (aquí entendida como tempo decorrido, em dias, até a ocorrência do sinistro) e tipo de perda. A seguir apresentaremos uma breve explicação de cada um delas.

- Valor de mercado do veículo: variável que caracteriza o valor do veículo no mercado. Sua referência de medida é a Tabela com preços de veículos desenvolvida

pela Fundação Instituto de Pesquisas Econômicas (FIPE). Esta tabela apura o valor médio de mercado dos veículos e é gerada com base em pesquisas realizadas em todo o País. Assume um valor médio de mercado para cada ano e modelo de veículo, nacional e/ou importado.

- Modelo do veículo: variável que representa a descrição do veículo. Em nosso estudo temos a grande maioria, mais de 95% dos casos, como veículos nacionais. A codificação que esta variável recebeu foi ordinal, tendo como parâmetro o seu valor de mercado. A seguir apresentamos tabela com as faixas de valores e os respectivos códigos.

Faixa de valores dos veículos	
Faixas	Códigos
Até R\$20.000,00	1
De R\$20.000,01 até R\$30.000,00	2
De R\$30.000,01 até R\$40.000,00	3
Acima de R\$40.000,00	4

- Ano do veículo: variável que representa o ano de fabricação do veículo. Nesta monografia, iremos utilizá-la como tempo de uso do veículo (idade), sendo seu valor obtido pela subtração de 2007 e o ano de fabricação.
- Tipo de franquia: conforme descrito na página 16, a franquia é a participação do segurado na ocorrência de sinistros de perda parcial. Sua caracterização é dada por agravada em 100%, agravada em 50%, agravada em 25%, básica, reduzida em 25% e reduzida em 50%. O valor de cada uma dessas caracterizações é mais elevado à medida que você vai reduzindo o agravamento, ou seja, o seguro que possui uma franquia agravada em 100% é mais barato que um que possui uma reduzida em 50%. No entanto, se houver a ocorrência do sinistro, o valor da franquia agravada em 100% é bem mais elevado que o reduzida em 50%. Neste estudo codificamos esta variável de forma categórica, sendo a categoria 1 referente a franquia básica, a 2 referente a franquia reduzida e a 3 referente a franquia agravada.
- Região de circulação: variável que caracteriza a região onde o segurado circula pela maior parte do tempo. Neste estudo trabalhamos com as regiões de Minas Gerais, São Paulo e Rio de Janeiro, onde aproximadamente 90% dos casos concentraram-se em Minas Gerais. Por este motivo, resolvemos classificar como demais regiões aquelas que fossem diferentes de Minas Gerais, ou seja, São Paulo e

Rio de Janeiro. A codificação desta variável foi categórica, assumindo o valor 1 para Minas Gerais e 0 caso contrário.

- Número de prestações: variável que caracteriza o número de parcelas que o segurado pagou o seguro. Sua codificação seguiu o critério de assumir o valor 0 se o seguro foi pago à vista ou, caso contrário, o número de prestações definidas na contratação do mesmo. Nos casos de sinistro de perda total, a definição foi dada pelo número de parcelas pagas até a data do evento.

- Código do corretor: variável que identifica o corretor dentro da companhia seguradora. Cada corretor tem um código próprio, onde através dele a seguradora realiza estudos de produtividade, riscos subscritos, condição de comercialização, dentre outros. Para codificar esta variável dividimos a base por região e dentro de cada região classificamos o corretor de forma ordinal com base em sua produtividade, ou seja, para um corretor de Minas Gerais que apresentou a menor produção dos casos em estudo, atribuímos o código 1. A apuração da produtividade foi dada pela quantidade de seguros vendidos por cada corretor. A seguir apresentamos Tabela ilustrativa dos sinistros não fraudulentos da região de São Paulo para o ano de 2005.

Codificação dos Corretores

Região - São Paulo / Ano – 2005		
Código Corretor	Seguros Vendidos	Codificação
61	13	4
4704	3	3
4702	2	2
9500	1	1
4400	1	1
3992	1	1
3357	1	1
2600	1	1
2202	1	1
1056	1	1
453	1	1
321	1	1
16	1	1

- Estado Civil: variável que caracteriza o estado civil do segurado. Neste estudo ela é categórica sendo que assume o valor 1 para solteiro e 0 para não solteiro.

- Garagem Comercial: variável que caracteriza se o segurado possui ou não garagem no trabalho. Geralmente esta informação não é obtida com frequência, uma vez que o seu preenchimento na contratação do seguro só é obrigatório quando o segurado

utiliza veículo para ir ao trabalho. Em nossa base de dados existem alguns registros caracterizados como não informado. Neste sentido, para utilização desta variável no modelo, adotamos como critério a avaliação se o veículo é usado para trabalho ou não. Com isso a variável foi codificada como categórica sendo 1 para veículo utilizado para trabalho e 0 se não utilizado.

- **Sexo:** variável que identifica o gênero do segurado tendo-se codificado com 0 para feminino e 1 para masculino.
- **Idade:** variável que representa a idade em anos do segurado quando da contratação do seguro.
- **Vigência:** variável que representa o número de dias decorridos desde a contratação do seguro até a ocorrência do sinistro. Sua codificação foi dada de forma categórica, assumido o valor 1 para sinistros prematuros e 0 para os não prematuros. Nos próximos parágrafos apresentamos uma explicação mais detalhada desta variável, bem como a definição de sinistro prematuro.
- **Tipo de perda:** o tipo de perda do veículo pode ser caracterizado como total ou parcial, sendo a classificação baseada no percentual de danos causados ao veículo. Quando o percentual é igual ou superior a 75% do limite máximo de indenização estipulado na apólice, o sinistro é caracterizado como de perda total. Logo, caso o percentual seja inferior, sua classificação é dada por parcial. Esta variável foi codificada de forma categórica, recebendo o valor 1 quando o sinistro for de perda total e valor 0 quando for de perda parcial.

Cada variável foi avaliada seguindo um critério simples de consistência, onde não era permitida a ausência de nenhuma informação ou a presença de informação sem relação com a realidade da variável. No caso da variável Garagem Comercial, a definição não informado vai de encontro com sua natureza, uma vez que, nos seguros contratados para veículos não utilizados para trabalho, esta variável é caracterizada automaticamente como não informada. Logo, entende-se que os registros definidos como não informado na base de dados são de seguros de veículos não utilizados para trabalho. Nesta avaliação não encontramos nenhum caso que fosse necessário ajuste, ou seja, todas variáveis estavam de acordo com sua natureza. Achamos pertinente comentar o tratamento especial que demos as variáveis vigência e modelo do veículo.

Na avaliação da variável vigência procuramos detectar sinistros prematuros, ou seja, sinistros ocorridos nas primeiras semanas – aqui definimos como sendo as 7 primeiras ou 50 dias - de contratação do seguro. Esses sinistros, conforme contato com os profissionais da seguradora, apresentam forte indício de fraude. Nos 186 casos de sinistros fraudulentos que conseguimos coletar, foram encontrados 43 registros com esse perfil, ou seja, 23,12%. Por

curiosidade, na amostra dos 600 casos de sinistros não fraudulentos, foram encontrados 295 casos, uma representatividade de 49,16%. Nestes casos, os sinistros podem ter sido classificados como não fraudulentos por motivo da falta de sistemas eficazes de comprovação das evidências de má fé do segurado, ou até mesmo não serem sinistros com intenção de fraude por parte do segurado.

Para a variável modelo do veículo, visando minimizar a quantidade de modelos do mesmo porte, agrupamos os mesmos conforme descrito a seguir. Por exemplo, o veículo Uno Mille aparecia diversas vezes na base apresentando os seguintes modelos:

- UNO MILLE EX 1.0 IE (G) 2P;
- UNO MILLE 1.0 8V FIRE 4P;
- UNO MILLE 1.0 8V FIRE 2P.

então, resolvemos padronizá-los para a categoria Uno.

7. Resultado da Pesquisa

Após ajustar o modelo de Regressão Logística utilizando o software Minitab versão 14, verificou-se várias situações para as diversas simulações que foram realizadas. Inicialmente foi ajustado o modelo logístico na base de dados completa, ou seja, com 786 observações sendo que destas 186 referiam-se a sinistros fraudulentos e 600 a sinistros não fraudulentos. Conforme apresentado na Tabela 1, as variáveis ano do veículo, aqui entendida como idade do veículo, região de circulação, código do corretor, sexo e valor de mercado, não eram significativas para o modelo. Isso se deu devido à probabilidade de significância dessas variáveis terem sido inferiores ao nível de significância nominal adotado neste estudo, que é de 5%.

Tabela 1: Resultado da aplicação do modelo de regressão logística completo

Variáveis	Coefficiente estimado (β)	Erro Padrão	Prob. de Significância	Odds
Mod. Veículo	-0,366644	0,166551	0,0280	0,6900
Ano Veículo	0,012204	0,035081	0,7280	1,0100
Tip. Franquia	-0,513325	0,209278	0,0140	0,6000
Reg. Circulação	0,524056	0,301279	0,0820	1,6900
Num. Prestações	0,533230	0,051732	0,0000	1,7000
Cód. Corretor	-0,025701	0,051285	0,6160	0,9700
Est. Civil	0,734414	0,290760	0,0120	2,0800
Veic. Trabalho	-2,196110	0,380995	0,0000	0,1100
Sexo	0,177554	0,247499	0,4730	1,1900
Idade (Segurado)	0,041736	0,011160	0,0000	1,0400
Tip. Perda	0,863463	0,243388	0,0000	2,3700
Vlr. Mercado	0,000007	0,000007	0,3320	1,0000
Vigência	-0,885714	0,246231	0,0000	0,4100
Constante	-1,889640	0,841851	0,0250	

Uma vez identificado as variáveis que seriam excluídas, gerou-se o novo modelo com aquelas que permaneceram. A seguir apresentamos a Tabela 2 com os resultados obtidos.

Tabela 2: Resultado do modelo de regressão logística incompleto

Variáveis	Coefficiente estimado (β)	Erro Padrão	Prob. de Significância	Odds
Mod. Veículo	-0,286922	0,103840	0,0060	0,7500
Tip. Franquia	-0,476651	0,204285	0,0200	0,6200
Num.	0,526438	0,050691	0,0000	1,6900
Est. Civil	0,766600	0,283693	0,0070	2,1500
Veic. Trabalho	-2,252660	0,380058	0,0000	0,1100
Idade	0,041986	0,010951	0,0000	1,0400
Tip. Perda	0,856880	0,233997	0,0000	2,3600
Vigência	-0,883215	0,244814	0,0000	0,4100
Constante	-1,344230	0,741893	0,0700	

De posse do novo resultado, foi criada a tabela de classificação usando o método da substituição para estimação das probabilidades de erro, e verificou-se que o modelo tem melhor capacidade preditiva para os sinistros fraudulentos do que para os sinistros não

fraudulentos, pois as porcentagens de classificação corretas são respectivamente 80,65% e 77,67%.

Tabela 3: Tabela de classificação

Grupo Real	Grupo Classificado pelo modelo			
	Fraudulentos	Não Fraudulentos	Total	% de Acerto
Fraudulentos	150	36	186	80,65%
Não Fraudulentos	134	466	600	77,67%

Conforme se observa na tabela acima, classificando a base de dados utilizada para gerar o modelo, a assertividade para os grupos dos fraudulentos e não fraudulentos foi de 80,65% e 77,67% respectivamente. Isso quer dizer que ao classificar a base que foi utilizada para gerar o modelo, um razoável percentual de erro de discriminação foi verificado.

Após avaliação do ajuste e da classificação do modelo gerado pela base completa, iniciou-se uma série de simulações com a amostra mais balanceada, ou seja, com 386 observações, sendo que destas 186 se referiam a sinistros fraudulentos e 200 a sinistros não fraudulentos. Para selecionar as amostras dos sinistros não fraudulentos, retirou-se aleatoriamente da amostra total de 600 observações, 200 sinistros de cada vez. Uma vez selecionada a primeira amostra, retornou-se com os 200 sinistros para base completa e uma nova amostra foi selecionada. Assim foi procedido até gerar a décima amostra.

Uma vez selecionadas as dez amostras procedeu-se conforme feito no modelo gerado com a base completa. Os modelos foram ajustados retirando as variáveis não significativas e depois foi gerado a tabela de classificação. A seguir apresentamos na Tabela 4 os valores resumos das classificações dos modelos gerados. Optamos por deixar a constante para verificar o efeito matemático da classificação, mesmo que em alguns casos o p-valor do modelo tenha sido elevado, maior que 5%.

Tabela 4: Classificação dos modelos simulados - com constante

Simulações	Fraudulentos	Não Fraudulentos	Total	% de Acerto	Tipo
Base Geral	92	94	186	49,46%	Fraudulentos
	30	570	600	95,00%	Não Fraudulentos
Simulação 1	142	44	186	76,34%	Fraudulentos
	33	167	200	83,50%	Não Fraudulentos
Simulação 2	142	44	186	76,34%	Fraudulentos
	34	166	200	83,00%	Não Fraudulentos
Simulação 3	145	41	186	77,96%	Fraudulentos
	33	167	200	83,50%	Não Fraudulentos
Simulação 4	144	42	186	77,42%	Fraudulentos
	33	167	200	83,50%	Não Fraudulentos
Simulação 5	145	41	186	77,96%	Fraudulentos
	40	160	200	80,00%	Não Fraudulentos
Simulação 6	142	44	186	76,34%	Fraudulentos
	34	166	200	83,00%	Não Fraudulentos
Simulação 7	151	35	186	81,18%	Fraudulentos
	32	168	200	84,00%	Não Fraudulentos
Simulação 8	142	44	186	76,34%	Fraudulentos
	32	168	200	84,00%	Não Fraudulentos
Simulação 9	143	43	186	76,88%	Fraudulentos
	33	167	200	83,50%	Não Fraudulentos
Simulação 10	141	45	186	75,81%	Fraudulentos
	35	165	200	82,50%	Não Fraudulentos

Na Tabela 5 apresentamos os mesmos modelos da Tabela 4, no entanto com os escores calculados sem a constante.

Tabela 5: Classificação dos modelos simulados - sem constante

Simulações	Fraudulentos	Não Fraudulentos	Total	% de Acerto	Tipo
Base Geral	150	36	186	80,65%	Fraudulentos
	134	466	600	77,67%	Não Fraudulentos
Simulação 1	172	14	186	92,47%	Fraudulentos
	72	128	200	64,00%	Não Fraudulentos
Simulação 2	141	45	186	75,81%	Fraudulentos
	34	166	200	83,00%	Não Fraudulentos
Simulação 3	156	30	186	83,87%	Fraudulentos
	46	154	200	77,00%	Não Fraudulentos
Simulação 4	178	8	186	95,70%	Fraudulentos
	94	106	200	53,00%	Não Fraudulentos
Simulação 5	173	13	186	93,01%	Fraudulentos
	73	127	200	63,50%	Não Fraudulentos
Simulação 6	160	26	186	86,02%	Fraudulentos
	48	152	200	76,00%	Não Fraudulentos
Simulação 7	179	7	186	96,24%	Fraudulentos
	82	118	200	59,00%	Não Fraudulentos
Simulação 8	155	31	186	83,33%	Fraudulentos
	42	158	200	79,00%	Não Fraudulentos
Simulação 9	128	58	186	68,82%	Fraudulentos
	15	185	200	92,50%	Não Fraudulentos
Simulação 10	183	3	186	98,39%	Fraudulentos
	122	78	200	39,00%	Não Fraudulentos

Nota-se que aparentemente é melhor ficar sem a constante para classificação dos sinistros. Observamos também que os modelos gerados com as amostras balanceadas, 386 observações, apresentam melhores resultados quanto à classificação dos sinistros fraudulentos. Outro ponto que achamos interessante de apresentar foi o que ocorreu em cada caso, ou seja, quais variáveis sempre se destacaram e quais as que foram sempre rejeitadas em cada simulação no ajuste do modelo logístico.

Tabela 6: Variáveis dos modelos simulados

Simulações	Variáveis que permaneceram	Variáveis rejeitadas
Base Geral	Modelo Veículo, Tipo Franquia, Número Prestações, Estado Civil, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Ano do Veículo - Região de Circulação - Código Corretor - Sexo - Valor de Mercado.
Simulação 1	Tipo de Franquia, Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Modelo Veículo - Ano do Veículo - Código Corretor - Estado Civil - Sexo - Valor de Mercado - Região de Circulação.
Simulação 2	Modelo Veículo, Tipo de Franquia, Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Valor de Mercado, Vigência.	Ano do Veículo - Região de Circulação - Código Corretor - Estado Civil - Sexo.
Simulação 3	Tipo de Franquia, Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Modelo Veículo - Ano do Veículo - Região de Circulação - Código Corretor - Estado Civil - Sexo - Valor de Mercado.
Simulação 4	Modelo Veículo, Região de Circulação, Número de Prestações, Estado Civil, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Ano do Veículo - Tipo de Franquia - Código Corretor - Sexo - Valor de Mercado.
Simulação 5	Modelo Veículo, Número de Prestações, Estado Civil, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Ano do Veículo - Tipo de Franquia - Código Corretor - Sexo - Valor de Mercado - Região de Circulação.
Simulação 6	Tipo de Franquia, Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Ano do Veículo - Região de Circulação - Código Corretor - Estado Civil - Sexo - Valor de Mercado - Modelo Veículo.
Simulação 7	Modelo Veículo, Tipo de Franquia, Região de Circulação, Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Ano do Veículo - Código Corretor - Estado Civil - Sexo - Valor de Mercado.
Simulação 8	Modelo Veículo, Tipo de Franquia, Região de Circulação, Número de Prestações, Estado Civil, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Ano do Veículo - Código Corretor - Sexo - Valor de Mercado.
Simulação 9	Modelo Veículo, Tipo de Franquia, Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Valor de Mercado, Vigência.	Ano do Veículo - Região de Circulação - Código Corretor - Estado Civil - Sexo.

Simulação	Número de Prestações, Veículo Utilizado para Trabalho, Idade, Tipo de Perda, Vigência.	Modelo Veículo - Ano do Veículo - Tipo de Franquia - Região de Circulação - Código Corretor - Estado Civil - Sexo - Valor de Mercado.
-----------	--	---

Conforme pode ser visto na tabela 6, as variáveis que sempre permaneceram nos modelos simulados são: veículo utilizado para trabalho, idade, tipo de perda, vigência. Aquelas que sempre foram rejeitadas são: ano do veículo, sexo e código do corretor.

7.1 Validação do Resultado

Após todo o trabalho de ajuste e classificação dos modelos simulados, o próximo passo foi selecionar um conjunto de observações totalmente independente daquele utilizado para geração dos modelos, para então validar a predição. Nesse conjunto de dados foram levantados a quantidade máxima de sinistros fraudulentos, uma vez que os registros da Cia Seguradora não apresentava grande frequência desses sinistros. Conseguiu-se então extrair 15 sinistros fraudulentos e por consequência foram extraídos 15 sinistros não fraudulentos para equilibrar a amostra de validação.

De posse da amostra de validação foram estimados os escores de cada modelo simulado, incluindo o gerado com a base completa sem a constante, e chegou-se na Tabela de classificação apresentada a seguir.

Tabela 7: Validação dos modelos simulados

Simulações	Fraudulentos	Não Fraudulentos	Total	% de Acerto	Tipo
Base Geral	5	10	15	33,33%	Fraudulentos
	2	13	15	86,67%	Não Fraudulentos
Simulação 1	14	1	15	93,33%	Fraudulentos
	5	10	15	66,67%	Não Fraudulentos
Simulação 2	12	3	15	80,00%	Fraudulentos
	2	13	15	86,67%	Não Fraudulentos
Simulação 3	13	2	15	86,67%	Fraudulentos
	4	11	15	73,33%	Não Fraudulentos
Simulação 4	9	6	15	60,00%	Fraudulentos
	3	12	15	80,00%	Não Fraudulentos
Simulação 5	14	1	15	93,33%	Fraudulentos
	5	10	15	66,67%	Não Fraudulentos
Simulação 6	14	1	15	93,33%	Fraudulentos
	4	11	15	73,33%	Não Fraudulentos
Simulação 7	14	1	15	93,33%	Fraudulentos
	7	8	15	53,33%	Não Fraudulentos
Simulação 8	13	2	15	86,67%	Fraudulentos
	4	11	15	73,33%	Não Fraudulentos
Simulação 9	9	6	15	60,00%	Fraudulentos
	2	13	15	86,67%	Não Fraudulentos
Simulação 10	10	5	15	66,67%	Fraudulentos
	4	11	15	73,33%	Não Fraudulentos

Com base nos resultados da Tabela 7, entendeu-se que o melhor modelo para este estudo seria o apresentado na simulação 6, uma vez que ele apresenta maior assertividade para o grupo dos sinistros fraudulentos e dentre aqueles que apresentaram o mesmo percentual de acerto (simulação 1, 5 e 7), ele é quem apresenta maior assertividade para os sinistros não fraudulentos. A seguir apresentamos na Tabela 8 as variáveis, os coeficientes, erro padrão, probabilidade de significância e razão de chance do modelo.

Tabela 8: Resultado do modelo logístico da simulação 6

Variáveis	Coefficiente estimado (β)	Erro Padrão	Prob. de Significância	Odds
Tip. Franquia	-0,850589	0,256739	0,0010	0,4300
Num. Prestações	0,596727	0,072864	0,0000	1,8200
Veic. Trabalho	-1,649610	0,506045	0,0010	0,1900
Idade	0,030867	0,012707	0,0150	1,0300
Tip. Perda	1,294160	0,320121	0,0000	3,6500
Vigência	-0,952488	0,301613	0,0020	0,3900
Constante	-0,415751	0,882938	0,6380	

A distribuição dos escores dos sinistros fraudulentos e não fraudulentos do modelo gerado com a base completa sem a constante e do modelo gerado na simulação 6 são apresentados nas figuras a seguir:

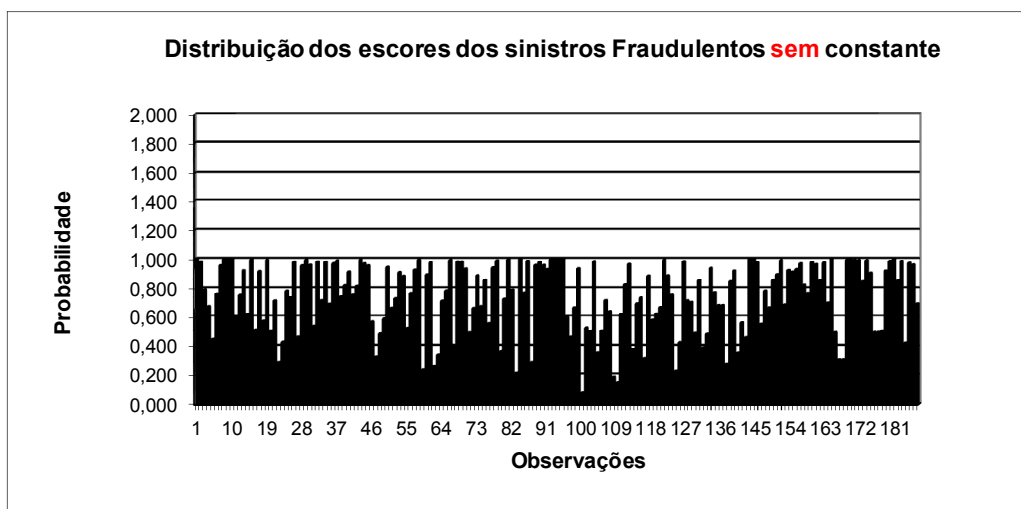


Figura 1 – Escores dos sinistros fraudulentos do modelo gerado com a base

Nota-se que para os sinistros fraudulentos, a grande maioria (80,65%) das barras está acima de 0,5 do eixo y, ou seja, caracterizando o sinistro como pertencente ao grupo dos fraudulentos.

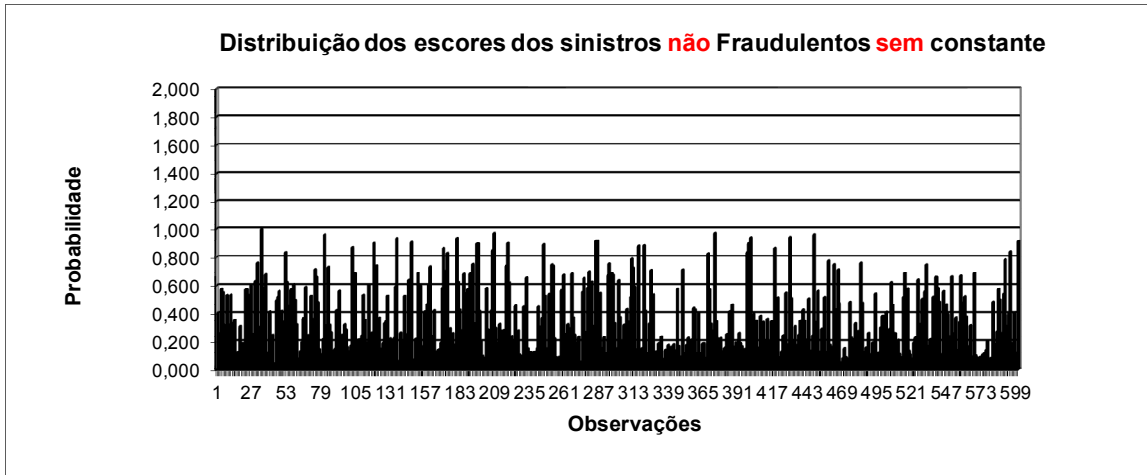


Figura 2 – Escores dos sinistros não fraudulentos do modelo gerado com a base

Já para os sinistros não fraudulentos, a grande maioria (77,67%) das barras está abaixo de 0,5 do eixo y, ou seja, caracterizando do sinistro como pertencente ao grupo dos não fraudulentos.

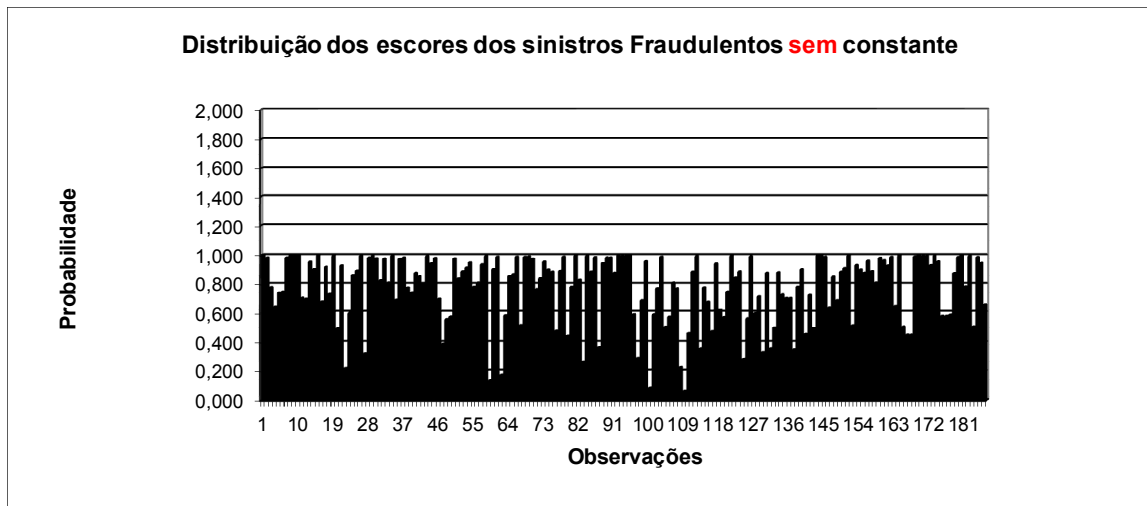


Figura 3 – Escores dos sinistros fraudulentos do modelo gerado na simulação 6.

Para os escores dos sinistros fraudulentos gerados na simulação 6, o comportamento das barras retratam uma situação ainda melhor quando comparado com o modelo gerado com a base completa, ou seja, 93,33% das barras estão acima de 0,5 do eixo y.

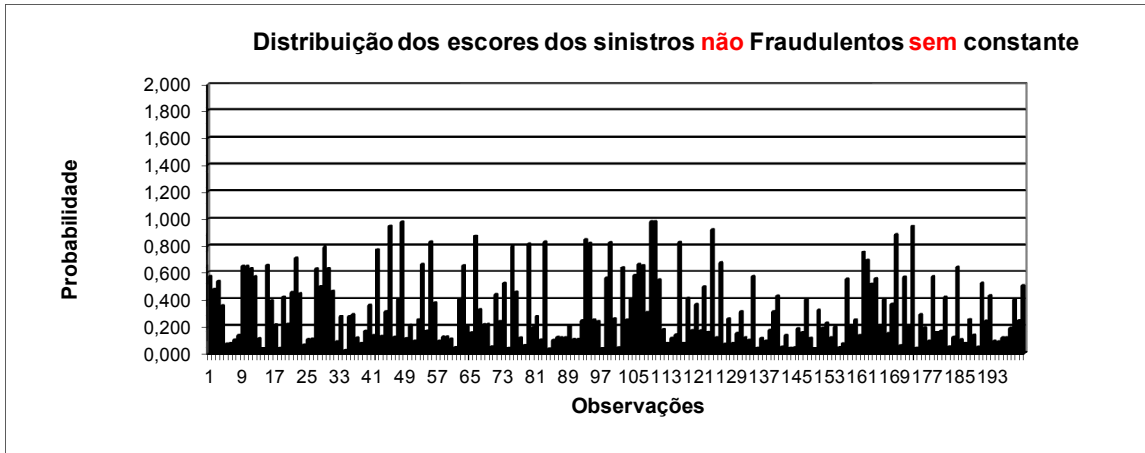


Figura 4 – Escores dos sinistros não fraudulentos do modelo gerado na simulação 6.

No entanto, a realidade verificada para os sinistros fraudulentos não se repete para os sinistros não fraudulentos. Nota-se aqui que 73,33% das barras estão abaixo de 0,5 do eixo y. Para o modelo gerado com a base completa observa-se uma assertividade de 77,67%.

A seguir apresentamos a formulação de discriminação do modelo gerado na simulação 6.

Grupo dos Fraudulentos:

$$\Pi(z) = \frac{\exp^{(-0,8506.(Tp\ Franq)+0,5967.(N^{\circ}\ Prest)-1,6496.(Veic\ Trab)+0,0309.(Idade)+1,2941.(Tp\ Perda) -0,9525.(Vig))}}{1 + \exp^{(-0,8506.(Tp\ Franq)+0,5967.(N^{\circ}\ Prest)-1,6496.(Veic\ Trab)+0,0309.(Idade)+1,2941.(Tp\ Perda) -0,9525.(Vig))}}$$

Para todo $z = 1$.

Grupo dos não Fraudulentos:

$$\Pi(z) = \frac{1}{1 + \exp^{(-0,8506.(Tp\ Franq)+0,5967.(N^{\circ}\ Prest)-1,6496.(Veic\ Trab)+0,0309.(Idade)+1,2941.(Tp\ Perda) -0,9525.(Vig))}}$$

Para todo $z = 0$.

Logo, para verificar se um determinado segurado pertence ao grupo dos fraudulentos ou dos não fraudulentos é só aplicar os valores das variáveis explicativas deste nas formulações acima e verificar as probabilidades estimadas. Aquela que for maior é a que o segurado pertencerá.

Exemplo de aplicação do modelo:

Uma seguradora de médio porte do mercado mineiro está buscando uma regra para discriminar seus riscos subscritos no ramo de seguro de automóvel. Para isso criou dois grupos:

- Grupo dos fraudulentos: Constituído de segurados que contratam o seguro com má fé;
- Grupo dos não fraudulentos: Constituídos de segurados que contratam o seguro com boa fé.

Supondo que o modelo para a discriminação seja o ajustado na seção 4.2, para um segurado com franquia básica (categoria (1)), pagando o seguro de 4 vezes, que não usa veículo para ir ao trabalho (categoria (0)), possui 40 anos, no último ano sofreu sinistro de perda parcial (categoria (0)) e o tempo entre a contratação do seguro e a ocorrência do sinistros seja de 200 dias (categoria (0)), obtemos as seguintes estimativas de probabilidade para os grupos dos fraudulentos e dos não fraudulentos:

$$\hat{p}(\text{fraude}) = \frac{\exp^{(-0,8506 \cdot (1) + 0,5967 \cdot (4) - 1,6496 \cdot (0) + 0,0309 \cdot (40) + 1,2941 \cdot (0) - 0,9525 \cdot (0))}}{1 + \exp^{(-0,8506 \cdot (1) + 0,5967 \cdot (4) - 1,6496 \cdot (0) + 0,0309 \cdot (40) + 1,2941 \cdot (0) - 0,9525 \cdot (0))}} = 94,10\%$$

$$\hat{p}(\text{não fraude}) = \frac{1}{1 + \exp^{(-0,8506 \cdot (1) + 0,5967 \cdot (4) - 1,6496 \cdot (0) + 0,0309 \cdot (40) + 1,2941 \cdot (0) - 0,9525 \cdot (0))}} = 5,90\%$$

Portanto, este segurado seria classificado como pertencendo ao grupo dos fraudulentos, uma vez que a respectiva probabilidade estimada é maior para este grupo que para o grupo dos não fraudulentos.

A seguir, apresentamos na Tabela 20 as faixas de escore para uma melhor classificação o risco.

Tabela 20: Faixas de classificação do risco

Score	Classe de Risco de fraude
0% a 24,99%	Não há risco de fraude
25% a 49,99%	Baixo risco de fraude
50% a 74,99%	Médio risco de fraude
75% a 100%	Alto risco de fraude

É importante ressaltar que, por ser um modelo de regressão, o modelo logístico sofre a influência da correlação entre variáveis explicativas. A presença de multicolinearidade afeta a confiabilidade das estimativas obtidas para os parâmetros do modelo (Mingoti, 2005).

8. Conclusão

Após a realização de todo o processo de modelagem, percebe-se que a fraude em sinistros de automóvel é influenciada por seis variáveis que caracterizam-se por tipo de franquia, número de prestações, veículo utilizado para trabalho, idade, tipo de perda e vigência. Acharmos interessante mencionar que um segurado classificado como fraudulento tem 3,65 vezes a mais de chance de cometer um sinistro de perda total do que um segurado classificado como não fraudulento. Esta situação evidencia a hipótese de que segurados que cometem fraude em seguros de automóvel, em grande parte das vezes, se encontram em situações financeiras delicadas. Vale ressaltar que devido à pequena amostra que colhemos de sinistros fraudulentos, essa conclusão tem suas limitações. Novos estudos podem ser realizados utilizando-se de base de sinistros fraudulentos com maior número de observações para então verificar as diversas situações encontradas.

Outro ponto que achamos relevante mencionar é sobre a variável corretor. Algumas companhias seguradoras entendem ser o corretor um grande aliado dos segurados quando da aplicação da fraude. Especialistas de regulação de sinistros também dizem que isso geralmente ocorre devido ao corretor conhecer bem a tramitação do seguro.

De fato entendemos que o mercado se prepara cada vez mais para evitar a fraude. Algumas seguradoras criam áreas específicas para regulação de sinistros especiais, a CNseg faz seu papel de quantificar e distribuir para o mercado a frequência das fraudes nos diversos ramos e eu proponho aqui, mais um mecanismo para reduzir a frequência da prática da fraude que prejudica a sociedade seguradora como um todo. O que esse mecanismo apresenta de

diferente do que hoje já existe é o fato de tentar detectar a fraude antes mesmo de ela ocorrer, ou seja, na subscrição do risco. Para isso, a companhia seguradora terá que gerar os modelos e fazer um trabalho de acompanhamento dos mesmos. Um trabalho bem feito nessa linha pode gerar grandes economias em longo prazo.

Como sugestão para novos estudos deixa-se aqui a ideia da geração dos modelos com uma flexibilidade maior quanto à seleção de variáveis que comporão o mesmo. Entende-se que uma probabilidade de significância de até 0,1 (ou 10%) pode gerar melhores estimativas de classificação, melhorando assim o resultado almejado. Outra possibilidade é categorizar as variáveis contínuas como idade. Dessa forma, pode-se chegar a modelos com uma capacidade preditiva melhor.

9. Referências

ALVIM, P. O Contrato de Seguros. 3 ed. Rio de Janeiro: Editora Forense, 1999. Cap. 8, p. 102-118 p.

MINGOTI, S .A. Análise de dados através de métodos de Estatística Multivariada. 1 ed. Minas Gerais. Editora UFMG, 2005. 295p.

ALBERTI, Verena et al. Entre a solidariedade e o risco: história do seguro privado no Brasil. 2 ed. Rio de Janeiro: Editora FGV, 2001. 352p.

BASTOS, Marcelo Lessa. O Ministério Público no Combate a Fraude nos Seguros. Disponível em: <<http://www.fdc.br/>>. Acesso em: 29/03/2006.

BRITES, José Carlos. Fraude em Seguros. Revista Bolsa dos Seguros, Rio de Janeiro, v.9, n.22, p.20-27, jan. 2006.

CAMPOS, José Elísio Ferraz de. Fraude em seguros. Revista de Seguros, Rio de Janeiro, v.86,n.852,p.8-11,jan/mar. 2005.

FALCÃO, Juliana. **Analista de Risco**. Disponível em: <<http://carreiras.empregos.com.br/comunidades/campus/profissoes/130701-atuaria.shtm>>. Acesso em:29/04/2006.

FERREIRA, Paulo Pereira. Modelos de precificação e ruína para seguros de curto prazo. 1 ed. Rio de Janeiro: FENASEG, 2002. Cap. 12, p. 185-195.

FERREIRA, Weber José. **Coleção Introdução à Ciência Atuarial**. 1 ed. Rio de Janeiro: 1985. 4v.

HAIR, Joseph F. et al. **Análise Multivariada de dados**. 5 ed. Porto Alegre: Editora Bookman, 2005. Cap. 1, p. 23-45.

HERSZKOWICZ, Fabio. **Credit Scoring**: A aplicação de métodos estatísticos na avaliação de risco de crédito. Tese (Doutorado) – Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade, São Paulo.

MARQUES, Lúcio Antônio. A doença que contaminou o mercado. **Revista Previdência e Seguros**. Rio de Janeiro, v.63, n.558, p.6-9, mar/abr. 2001.

MENDES, João José de Souza. **Bases Técnicas de Seguros**. 1 ed. São Paulo: Manuais Técnicos de Seguros, 1997. Cap. 1. p. 11-14.

NEGRINI, Pedro Paulo. A arte de gerenciar crimes inteligentes. **Revista Rumos**. Rio de Janeiro, v.25, n.183, p.8-12, abr. 2001.

SUSEP, Superintendência de Seguros Privados. **Termos Técnicos de Seguros**. Disponível em: <www.susep.gov.br>. Acesso em: 11/05/2013.

CNseg, Confederação Nacional das Empresas de Seguros Gerais, Previdência Privada e Vida, Saúde Suplementar de Capitalização. **Quantificação da Fraude no mercado de seguros brasileiro**. Disponível em: <www.fenaseg.org.br>. Acesso em 11/05/2013.

TZIRULNIK, Ernesto. Denúncias de fraude preocupam o mercado. **Revista Apólice**. São Paulo, v.10, n.84, p.12-15, mai. 2005.

XAVIER, Wagner Attina. **Introdução ao seguro**. 2 ed. Rio de Janeiro: FENASEG 1999. 62 p.