

## 4 Estudo empírico

Este capítulo apresentará o comportamento da metodologia, diante de dados empíricos. O capítulo se divide em 3 seções.

Na primeira seção, serão apresentadas as características gerais da base de dados empíricos utilizada na aplicação da metodologia.

Na segunda seção, serão apresentadas as características dos contratos com mais detalhes. Ainda nesta seção, uma proposta de reformulação das variáveis de perfil dos contratos, será apresentada. A reformulação será realizada com base na distribuição da frequência de sinistros por contrato, e na distribuição de apólices vigentes.

Finalmente, na terceira seção serão apresentados os agrupamentos das variáveis que compõe a classificação de risco. Serão reportados detalhes sobre o número de informações por classe, bem como um breve histórico das iterações do algoritmo. Toda a análise será realizada, considerando os agrupamentos formados, diante dos critérios sugeridos no capítulo 3.

### 4.1. Base de dados

As bases de dados atuariais das companhias de seguros brasileiras seguem um padrão que se assemelha, em muito, ao padrão sugerido por Westenberger e Teixeira (1991). Cabe ressaltar que, embora aquele trabalho tenha sido desenvolvido no início da década de 90, os problemas abordados ainda são encontrados atualmente.

A base de dados utilizada no presente estudo foi formada por uma amostra aleatória de 871.322 apólices que formam a carteira de algumas seguradoras brasileiras. Tais apólices iniciaram suas vigências entre Março/2002 e Fevereiro/2003, e totalizaram 428.008 itens expostos.

Além das informações referentes à vigência das apólices e aos contratos, a base contém informações de sinistros ocorridos entre Março/2002 e Fevereiro/2003. As indenizações destes sinistros são consideradas de duas formas

distintas: as indenizações pagas entre Março/2002 e Fevereiro/2003, e as indenizações pendentes até Maio/2003.

A base contempla 44.078 sinistros, que totalizam um volume de indenizações em torno dos 220 milhões de reais.

A estrutura de classificação de risco contida na base é formada pela combinação de três variáveis de identificação: tipos de veículos, regiões de tarifação, e idade dos veículos. Serão considerados, ao todo, 164 tipos de veículos, 50 regiões de tarifação e 12 idades. Com isso, obtém-se um total de 90.200 classes distintas, onde nem todas apresentam itens expostos. Durante o período contemplado pela base, apenas 34.475 classes de risco apresentaram informações, o que representa 38% do total de classes possíveis.

Na seção seguinte serão apresentados o comportamento e a distribuição da frequência de sinistros e apólices, diante das variáveis de perfil dos contratos. Após a análise, serão propostas reformulações daquelas variáveis, de forma a proporcionarem uma melhor visão do perfil de risco.

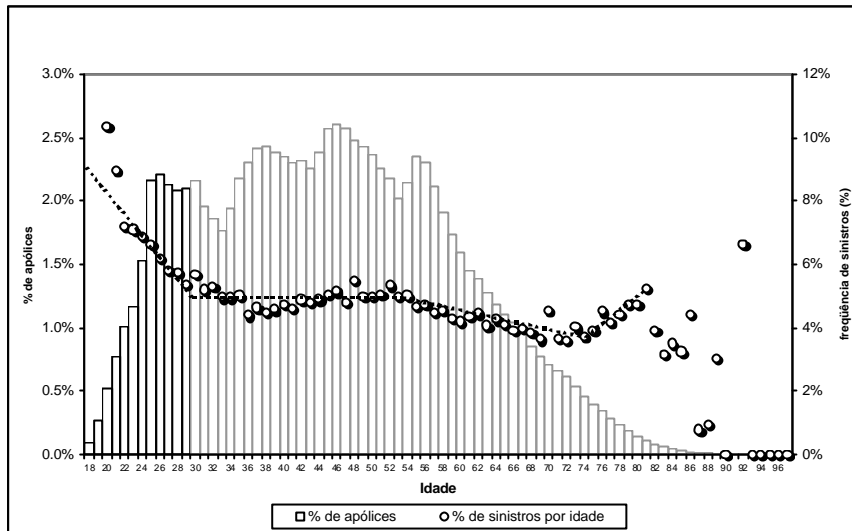
#### **4.2. Análise e reformulação das variáveis de perfil**

Conforme apresentado na seção 2.1.1, fazem parte deste estudo 11 variáveis de perfil, das quais 9 serão consideradas nesta seção.

Como a intenção deste trabalho é classificar o risco, as variáveis de perfil sofrerão algumas modificações baseadas na frequência de sinistros, que será expressa como o número médio de sinistros por apólice.

Inicialmente serão analisadas as distribuições de apólices e frequência de sinistros, dada a idade do condutor principal.

Figura 6: Distribuição de apólices e frequência de sinistros segundo a idade do condutor.



Na figura 6, as barras verticais representam a distribuição de apólices por idade, e o eixo vertical à esquerda, norteia seus valores. Nesta mesma figura, os pontos representam a frequência de sinistros, e seus valores encontram-se representados no eixo vertical à direita. A linha tracejada representa a tendência da frequência de sinistros com o aumento da idade dos condutores<sup>26</sup>.

A figura 6 corrobora as aspirações do mercado quanto à idade do condutor principal, pois mostra que a frequência de sinistros decresce com a idade, até uma idade limite.

As barras verticais margeadas em preto representam apólices onde os condutores se declararam com idade entre 18 e 29 anos. Como pode ser visto através da figura, a frequência de sinistros nesta faixa etária apresenta-se maior do que nas demais faixas. Além deste fato, nota-se que nesta faixa, com o aumento da idade, aumenta a representatividade das apólices diante do todo. Supõe-se que este comportamento decorra do poder aquisitivo dos jovens, associado aos valores de prêmios pouco convidativos geralmente praticados pelo mercado.

Observa-se que entre 30 e 54 anos a frequência de sinistros assume um comportamento aproximadamente constante. O mesmo ocorre com a distribuição de apólices, que se encontra representada pelas barras verticais margeadas em cinza. Nesta faixa, embora a distribuição das apólices apresente uma leve

<sup>26</sup> As figuras contidas nesta seção seguirão este mesmo padrão.

tendência, a mesma pode ser considerada estável, se comparada ao comportamento das demais faixas de idade.

Entre 55 e 74 anos, tanto a frequência de sinistros quanto o percentual de apólices decrescem. Suspeita-se que o comportamento decrescente da frequência tenha relação não só com o aumento da idade, como também com a responsabilidade familiar destes condutores, que geralmente são pessoas de referência em suas famílias.

Entre 75 e 79 anos, a frequência de sinistros assume um comportamento que contradiz a expectativa, pois não se acredita que pessoas nesta faixa de idade apresentem uma frequência de sinistros semelhante à de jovens. Suspeita-se que este comportamento seja reflexo de fraudes na declaração de condutores, que por sua vez são estimuladas pelo baixo controle de informações, associado ao valor convidativo dos descontos concedidos a condutores idosos.

A partir de 80 anos, a frequência de sinistros apresenta comportamento aleatório. Se for analisada em conjunto com o percentual de apólices, verifica-se que a representatividade das mesmas diante do todo é muito pequena.

Com base nestas informações, propõe-se a criação de três variáveis capazes de identificar condutores jovens (18 a 29 anos), adultos (30 a 54 anos), e idosos (55 anos ou mais), em detrimento do comportamento apresentado por condutores entre 75 e 79 anos de idade, já que se acredita que este comportamento não represente a realidade. Os intervalos de identificação propostos encontram-se diferenciados na figura 6.

Dando continuidade a análise das variáveis de perfil, a tabela abaixo mostra a distribuição de apólices e a frequência de sinistros segundo o sexo do condutor.

Tabela 2: Distribuição de apólices e frequência de sinistros segundo o sexo do condutor.

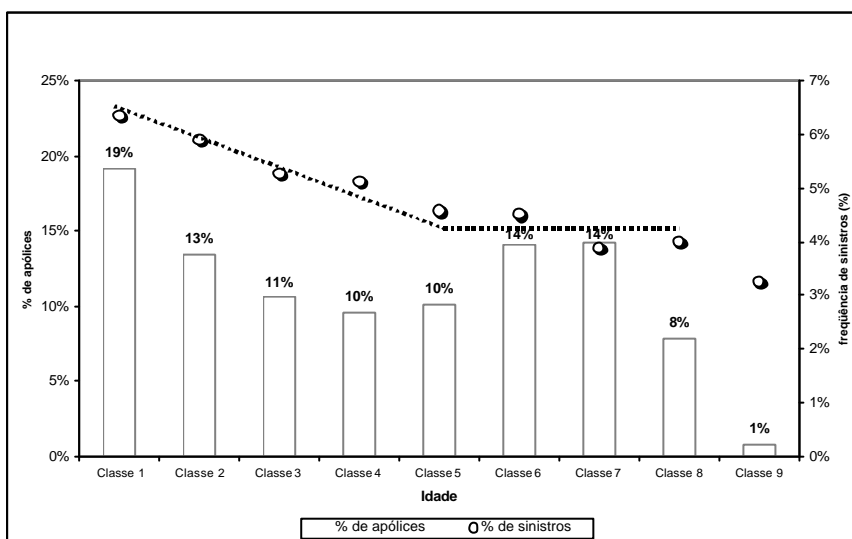
Sexo do Condutor	Percentual de Apólices	Frequência de Sinistros
Feminino	38.7%	4.7%
Masculino	61.3%	5.3%

Observando os dados reportados na tabela 2, nota-se que 38.7% dos condutores são mulheres, e que o percentual de homens é 58.4% maior. Observa-se ainda que homens apresentam frequência de sinistros 10.6% maior do que

mulheres. Supõe-se que o baixo percentual de condutores mulheres tenha relação com a discriminação por sexo, ainda presente na sociedade brasileira, e que, em detrimento do mito popular, mulheres sejam mais cautelosas do que homens, no que diz respeito a automóveis.

Considerando a figura 7 abaixo, que reporta o comportamento de sinistros e apólices segundo a classe de bônus, verifica-se que o percentual de apólices apresenta tendência decrescente. Apenas as classes 6 e 7 destoam do comportamento geral, apresentando percentual de apólices semelhante ao da classe 2.

Figura 7: Distribuição de apólices e freqüência de sinistros segundo a bonificação dos segurados.



Suspeita-se de duas hipóteses que possivelmente levariam as classes 6 e 7 a desviarem do comportamento de tendência esperado: bonificação falsa por parte dos canais de venda e descontos concedidos vantajosos em relação ao mercado.

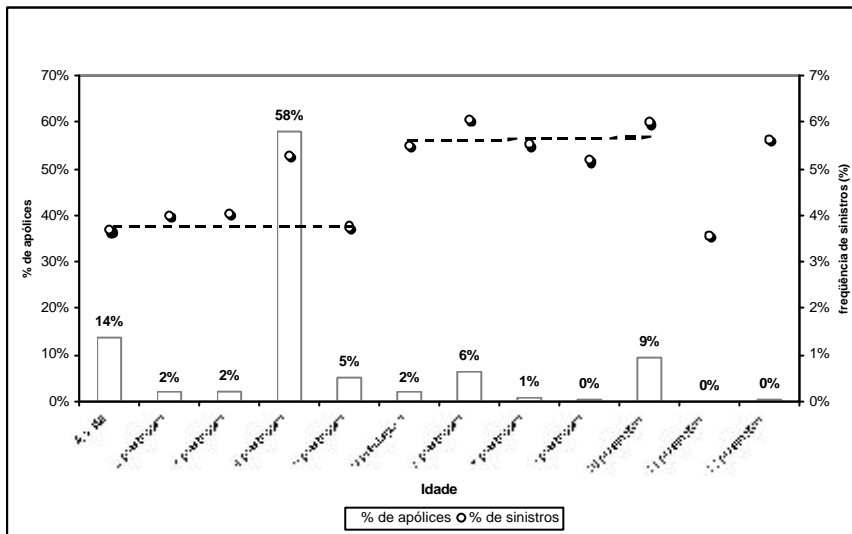
O comportamento dos dados, representado pela figura 7, sugere que haja correlação entre a freqüência de sinistros e a classe de bônus. Embora o fato não tenha sido comprovado estatisticamente, corrobora a própria definição de bonificação.

Analisando o comportamento da freqüência de sinistros, e desconsiderando a classe de bônus 9, devido ao baixo percentual de apólices, nota-se que, da classe 1 à classe 4, a freqüência de sinistros decresce com o aumento da bonificação, e que, a partir da classe de bônus 5, a freqüência apresenta comportamento

aproximadamente constante. Sabe-se ainda que o percentual de descontos concedido a classes de bônus acima de 4, são semelhantes, e por mais este motivo, decidiu-se por criar duas variáveis, uma capaz de identificar o maior risco, que é associado às classes de bônus de 1 a 4, e outra que identificasse o menor risco, representada pelas classes que vão de 5 a 9.

Em seguida, conforme mostra a figura 8 abaixo, será analisado o comportamento de sinistros e apólices segundo a forma de pagamento do prêmio.

Figura 8: Distribuição de apólices e freqüência de sinistros segundo a forma de pagamento do prêmio.



Observando a figura acima, que reporta o comportamento dos dados segundo a forma de pagamento, verifica-se que aproximadamente 58% dos contratos são quitados em 4 prestações. Suspeita-se que este fato decorra das atrativas taxas de juros cobradas.

Contratos pagos à vista (14%), e pagos em 10 prestações (9%), totalizam 23% do total das apólices. Suspeita-se que o comportamento das apólices com pagamento à vista, reflita o perfil de pessoas com melhores condições econômicas, ou mesmo, pessoas com perfil mais conservador. As demais formas de fracionamento do prêmio, que contabilizam aproximadamente 19% do total, encontram-se diluídas entre as demais opções de quitação.

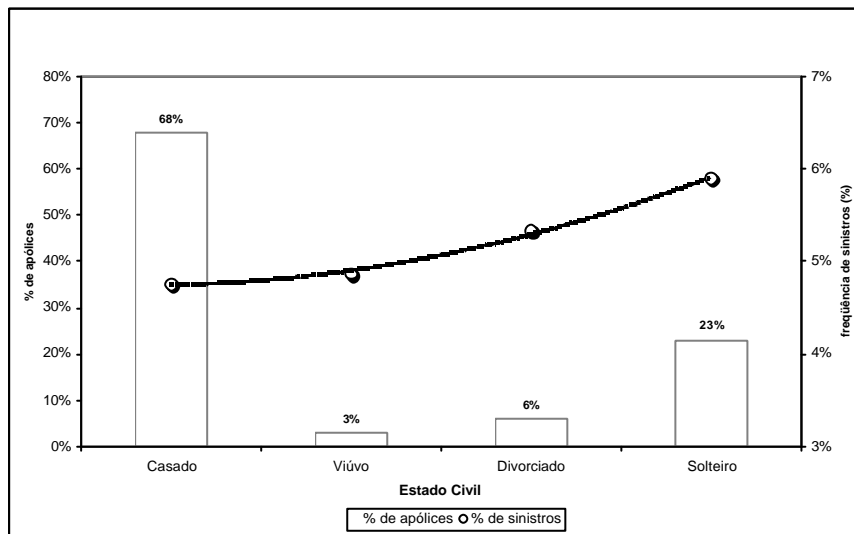
Desconsiderando os pagamentos em 11 e 12 parcelas, já que estes apresentam comportamento aleatório e baixa representatividade em relação ao

todo, é possível observar que a frequência de sinistros encontra-se em dois patamares distintos. Quando o prêmio é pago em até 5 parcelas, a frequência média de sinistros é de 3.7%, enquanto que a partir de 6 parcelas a média atinge aproximadamente o valor 5.6%.

Diante do exposto, e por acreditar que a forma de pagamento do prêmio possa representar as condições econômicas dos segurados, decidiu-se criar três variáveis capazes de identificar o comportamento observado: a variável que identifica quitações à vista, devido ao interesse particular por esta forma de pagamento, a variável que identifica quitações entre 2 e 5 parcelas e, finalmente, a variável que identifica quando o pagamento do prêmio foi realizado em 6 parcelas ou mais.

Em seguida serão analisados o percentual de apólices e a frequência de sinistros, segundo o estado civil dos condutores. Os dados encontram-se representados na figura abaixo.

Figura 9: Distribuição de apólices e frequência de sinistros segundo o estado civil do condutor principal.



De acordo com a figura, é possível verificar que o percentual de casados é de aproximadamente 68%. Condutores com esta característica, em detrimento dos demais, apresentam menor frequência de sinistros, o que pode ser interpretado como o amadurecimento e a responsabilidade dos cônjuges.

O número elevado de pessoas casadas pode ser resultante de duas hipóteses. Uma delas seria que, no período de análise, a falta de instrumentos legais que

subsidiassem ações efetivas no controle de informações, poderia ter levado muitos segurados a declarassem-se casados, já que os descontos concedidos a este estado civil são os maiores. Outra hipótese seria o maior interesse de pessoas casadas por seguros.

Analisando os demais estados, pode-se verificar que, com a redução da responsabilidade familiar do condutor, a frequência de sinistros aumenta. Observa-se ainda que condutores solteiros apresentam frequência de sinistros 23% maior do que a média dos demais. Com base nesta observação, decidiu-se criar duas variáveis de identificação, uma capaz de identificar condutores solteiros, e outra capaz de identificar os demais estados.

A tabela abaixo apresenta os dados referentes ao tipo de renovação.

Tabela 3: Distribuição de apólices e frequência de sinistros segundo o tipo de renovação.

Tipos de Renovação	Percentual de Apólices	Frequência de Sinistros
Seguro Novo	14.3%	6.7%
Congênere	16.3%	4.7%
Própria	69.4%	4.8%

Observando os dados, nota-se que 69.4% das renovações são da própria companhia. Suspeita-se que este fato reflita a satisfação e a fidelidade dos segurados, ou mesmo a efetividade da estratégia de preços adotada pela companhia. As 30.6% apólices restantes diluem-se entre renovações de congêneres e novos contratos de seguro.

Os dados mostram ainda que renovações próprias e de congêneres, apresentam aproximadamente, a mesma frequência de sinistros. Suspeita-se que este fato seja um indício da influência da classe de bônus e, conseqüentemente, da diminuição do risco com as renovações.

Devido ao desconhecimento do histórico de sinistros, é comum que seguros novos não recebam bonificação. A frequência de sinistros de 6.7%, 40% maior do que a frequência das demais renovações representaria uma justificativa para este fato.



Dado que cada tipo de renovação apresenta características distintas e que há um interesse particular no comportamento de cada tipo, foram criadas três variáveis de identificação.

A tabela 4 abaixo mostra a distribuição de apólices e a frequência de sinistros, segundo o tipo de franquia.

Tabela 4: Distribuição de apólices e frequência de sinistros segundo o tipo de franquia.

Tipos de Franquia	Percentual de Apólices	Frequência de Sinistros
Obrigatória	50.7%	4.5%
Reduzida	49.3%	5.6%

Analisando os dados, verifica-se que a distribuição das apólices é aproximadamente uniforme.

Observa-se ainda que a frequência de sinistros é 25% maior para contratos com franquia reduzida.

Sabendo que caso os reparos não atinjam o valor da franquia, o segurado se responsabilizará pelos danos, a franquia pode ser entendida com um dispositivo de exclusão de sinistros de baixo valor. Estes tipos de sinistros são representados pelas colisões cotidianas, e conseqüentemente, como sinistros de maior probabilidade de ocorrência. A franquia reduzida, conforme sugere o nome, representa um filtro menos seletivo de sinistros de baixo valor, o que explicaria a diferença percentual da frequência de sinistros entre a franquia obrigatória e a franquia reduzida.

Como geralmente o prêmio assume menor valor em casos onde a franquia obrigatória é contratada, acredita-se que esta variável seja capaz de representar, conjugada com outras, as condições econômicas dos segurados. Portanto, supondo esta premissa e diante do que pôde ser observado, foi criada uma variável identificadora de contratos cuja franquia contratada é obrigatória.

A seguir, na tabela 5, encontra-se reportado o comportamento dos dados segundo o tipo de indenização contratado.

Tabela 5: Distribuição dos dados segundo o tipo de indenização.

Formas de Indenização	Percentual de Apólices	Frequência de Sinistros
Valor Determinado	48.1%	5.3%
Valor de Mercado	51.9%	4.8%

Os dados contidos na tabela 5 mostram que as apólices se distribuem de forma quase homo gênea entre os níveis, diferindo apenas 4 pontos percentuais.

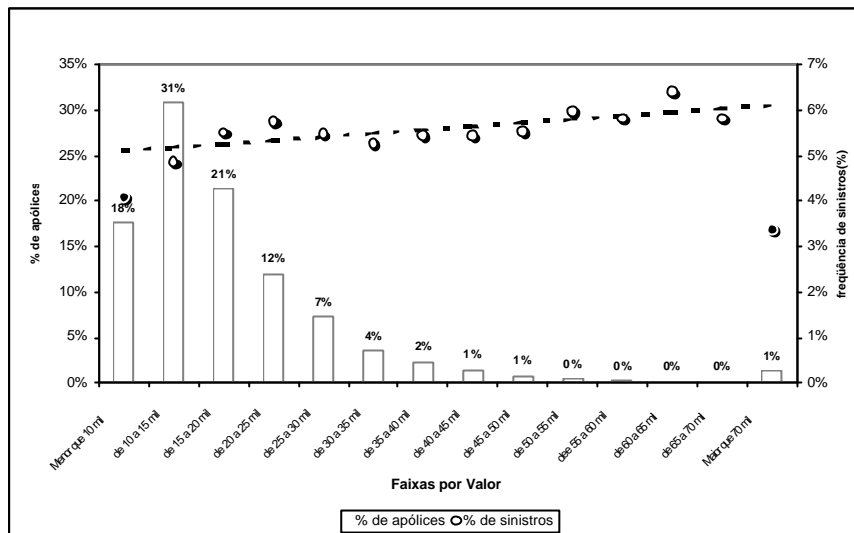
Este fato contrapõe-se à lógica, pois contratações de indenização por valor determinado deveriam tornar-se menos atrativas, devido aos agravos incidentes sobre o prêmio.

Algumas hipóteses combinadas podem ser responsáveis pelo aumento no número de contratos por valor determinado: a atratividade da política de descontos da companhia diante do mercado, a expectativa dos segurados pela desvalorização de seus veículos, e finalmente, sinistros intencionais. Sob a hipótese de que os segurados tenham acesso a estas informações, claramente a contratação de indenização através de valor determinado se apresenta mais atrativa, se comparada à indenização por valor de mercado. Diante do fato de que frequência de sinistros para indenizações com base em valor determinado é 10.8% maior do que as com base no valor de mercado, supõe-se que a hipótese de sinistros intencionais, que caracterizam irregularidades, teria maior participação no percentual de apólices.

Diante das informações apresentadas, extraiu-se o atributo que identifica reposição por valor de mercado.

Seria interessante observar o comportamento das apólices e dos sinistros, pelo valor da importância segurada. Para facilitar a interpretação dos dados, propôs-se a divisão da importância segurada, denominada I.S, em 14 faixas. Os resultados encontram-se reportados na figura abaixo.

Figura 10: Distribuição de apólices e freqüência de sinistros segundo faixas da I.S.



Conforme esperado, nota-se que existe um número menor de apólices com I.S. inferior a R\$ 10.000,00. A partir deste ponto, verifica-se que o percentual decresce com o aumento do valor da I.S., representando a própria situação econômica dos brasileiros.

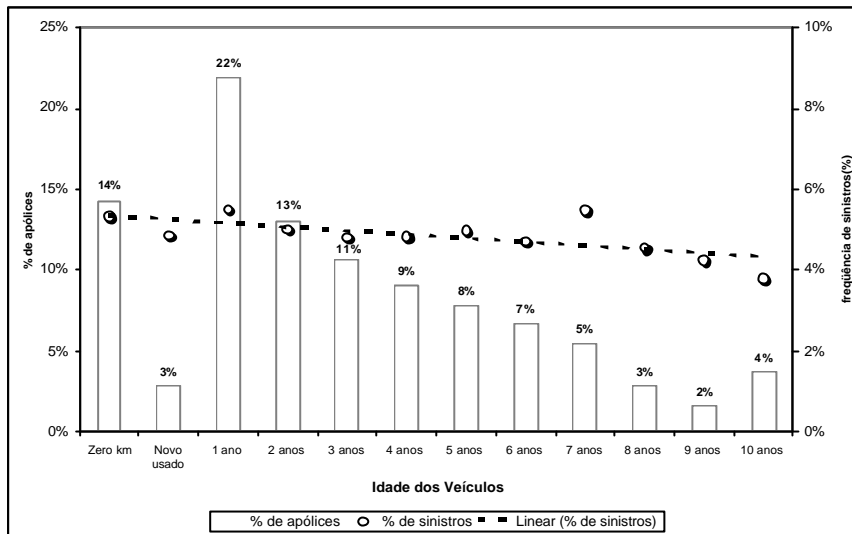
Observa-se ainda, que excluídos os dois extremos, a freqüência de sinistros cresce com o aumento do valor da I.S. Acredita-se que este fato tenha relação com veículos esportivos, ou mesmo com o descaso dos segurados com melhor situação econômica, por seus bens.

Sabe-se que algumas seguradoras, tanto nacionais quanto internacionais, adotam uma classificação de risco baseada na I.S., diferentemente da classificação analisada neste estudo. Por este motivo esta variável será utilizada como atributo na classificação dos automóveis.

A idade dos veículos é outra variável que compõe a classificação de risco na maioria das seguradoras.

Acredita-se que as sucessivas alterações de projetos de um veículo poderiam alterar suas características, e conseqüentemente, seu risco. Como a intenção da construção de atributos é utilizá-los para a classificação de veículos, será analisado o comportamento das apólices e dos sinistros segundo a idade.

Figura 11: Distribuição de apólices e frequência de sinistros segundo a idade dos veículos.



Observando a figura acima, nota-se que, com o passar dos anos, a frequência de sinistros decresce lentamente. Além disso, verifica-se que a distribuição das apólices, a despeito dos veículos novos usados<sup>27</sup>, apresenta um comportamento bem definido, e de acordo com o desejado pelas companhias<sup>28</sup>.

Ao final desta análise, foram construídos 18 atributos, conforme mostra a tabela abaixo.

<sup>27</sup> São considerados veículos novos usados, aqueles fabricados no ano da análise das informações, porém que foram usados entre a saída da concessionária e a contratação do seguro. Algumas seguradoras são flexíveis em relação ao período de uso dos veículos até decidirem classificá-los desta maneira.

<sup>28</sup> Embora os dados reportados pela figura 11 não demonstrem as razões, é comum que as companhias de seguros prefiram compor uma carteira jovem. Supõe-se que isto se deva à dificuldade de reparar veículos antigos, ao possível estado de conservação dos mesmos, e aos prêmios em média superiores, de veículos mais novos.

Tabela 6: Nomenclatura, definição e origem dos atributos.

Atributo	Definição	Variável Origem
Jovens	condutores entre 18 e 29 anos	idade do condutor
Adultos	condutores entre 30 e 54 anos	
Idosos	condutores com 55 anos e mais	
Sexo masculino	condutores do sexo masculino	sexo do condutor
Baixo risco	classes de bônus de 1 a 4	classe de bônus
Alto risco	classes de bônus de 5 a 9	
Quitação à vista	pagamentos à vista	forma de pagamento
Quitação de 2 a 5 parcelas	pagamentos entre 2 e 5 prestações	
Quitação em 6 parcelas ou mais	pagamentos em mais de 6 prestações	
Condutores solteiros	condutor solteiro	estado civil do condutor
Demais estados	condutores casados, divorciados ou viúvos	
Seguro novo	seguros cujo histórico de sinistros é desconhecido	tipo de renovação
Renovação congênere	renovações de seguros de origem de outras companhias	
Renovação própria	renovações de seguros da própria companhia	
Franquia obrigatória	franquia obrigatória	tipo de franquia
Valor de mercado	indenização com base no valor de mercado do veículo	forma de indenização
Importância segurada	importância segurada	importância segurada
Idade dos veículos	veículos entre zero quilômetro e de 10 anos ou mais	idade dos veículos

### 4.3. Aplicação metodológica

De acordo com o exposto na seção 3.1, a aplicação da metodologia segue a seqüência de passos do algoritmo. Por este motivo esta seção se dividirá em duas subseções.

Na primeira subseção, serão apresentados os agrupamentos iniciais formados por tipos de veículos e regiões de tarifação. Embora o algoritmo inicie com as idades desagregadas, a lógica de agrupamento das mesmas será apresentada ainda nesta subseção.

Na segunda subseção serão apresentados os agrupamentos finais formados diante de cada uma das três possibilidades de metas testadas.

### 4.3.1. Agrupamentos iniciais

Como estratégia de formação dos agrupamentos iniciais de veículos a proposta é a criação de fatores que sejam capazes de explicar grande parte da variância dos atributos<sup>29</sup>. Construídos os fatores, pretende-se utilizá-los para representar cada um dos 164 tipos de veículos, e então construir os agrupamentos.

Cabe ressaltar que os atributos, quando agregados para cada tipo de veículo, representam a proporção de apólices que se apresentam em cada estado. Com isso, os atributos proporcionais seriam capazes de informar, por exemplo, se dado veículo apresenta maior proporção de condutores jovens, em detrimento de adultos e idosos, ou mesmo se um veículo apresenta maior proporção de renovações de congêneres. A idéia por trás da construção dos fatores a partir dos atributos selecionados é permitir a representação do perfil dos segurados e condutores principais relacionando a própria preferência de cada um por certos tipos de veículos.

Conforme apresentado na seção 3.1, os fatores serão obtidos através da aplicação da técnica de análise fatorial. Esta técnica é utilizada principalmente quando a intenção é reduzir o número de atributos, sendo capaz de explicar grande parte da variância, e possibilitando a compreensão das relações existentes entre eles.

Existem várias técnicas utilizadas para estimar os carregamentos fatoriais e, conseqüentemente, a proporção da variância explicada pelos fatores. Como os atributos proporcionais formam uma matriz não singular, decidiu-se por aplicar o método das componentes principais na estimação dos carregamentos fatoriais.

Na aplicação da técnica de análise fatorial é necessário que seja definido, a priori, o número de fatores para os quais os escores fatoriais serão calculados, e que, conseqüentemente, substituirão os atributos.

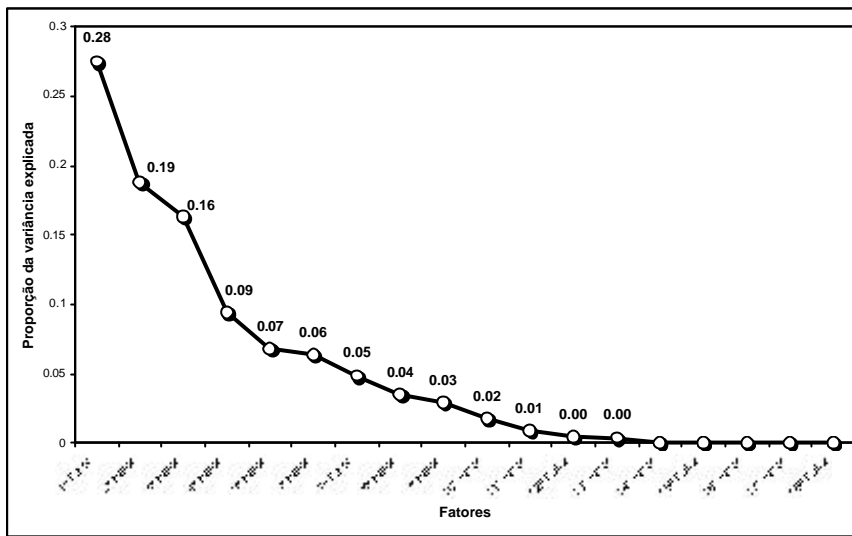
Várias são as ferramentas capazes de dar suporte à definição do número de fatores. Optou-se pela proporção da variância explicada por cada fator. Através desta medida procura-se investigar o ganho de explicação atribuído a cada fator. Como a representatividade do primeiro fator é superior à do segundo, e assim

---

<sup>29</sup> A lista dos atributos considerados pode ser vista na seção 4.2.

sucessivamente, é possível estimar a relevância dos fatores na explicação da variância.

Figura 12: Proporção da variância explicada através dos fatores.



Como mostra a figura acima, a partir do quarto fator, decai a representatividade dos autovalores, e, por isso, consideramos irrelevante a variância explicada pelos fatores posteriores - nível de relevância foi aqui estabelecido pelo decisor diante dos dados na tabela acima.

Os escores fatoriais (para os quatro primeiros fatores) foram estimados através do método de regressão<sup>30</sup>.

Devido à não singularidade da matriz de atributos, torna-se inviável a aplicação de testes estatísticos que comprovem a consistência dos fatores estimados. Por esta razão, optou-se por avaliar a consistência dos fatores através da estimação por partes<sup>31</sup>. Esta estratégia propõe que o universo de apólices seja dividido em duas amostras aleatórias, sem reposição, e de mesmo tamanho, e que então a análise fatorial seja aplicada em ambas amostras. Neste caso é importante que os carregamentos dos atributos dentro de cada fator, e para cada uma das amostras, representem a mesma relação<sup>32</sup>. De acordo com os resultados obtidos, a estimação se mostrou consistente.

<sup>30</sup> Os coeficientes das regressões que estimam os escores fatoriais constam no anexo 1.

<sup>31</sup> Johnson e Wichern (1998)

<sup>32</sup> Os resultados da análise de consistência dos fatores encontram-se no anexo 1.

Além dos escores fatoriais estimados, seria interessante conhecer a relação existente entre os atributos. Acredita-se que a partir deste conhecimento seja possível deduzir quais características levarão os veículos a se aglomerarem. Para ter acesso a esta informação é necessário conhecer os carregamentos dos atributos sobre cada um dos fatores. Os maiores carregamentos em cada fator apontarão quais os atributos que mais se correlacionam. Como nem sempre é possível detectar a relevância dos atributos em cada fator, foi aplicado o método de rotação Varimax.

Tabela 7: Carregamentos fatoriais e interpretação para os fatores.

Atributos	Fator 1	Fator 2	Fator 3	Fator 4
	Estrutura Familiar	Estratégia e Risco	Situação Financeira	Sexo do Condutor
<i>Faixa Etária</i>				
Jovens	<b>0.96</b>	-0.08	-0.03	0.08
Adultos	-0.28	-0.14	<b>-0.53</b>	-0.18
Idosos	<b>-0.58</b>	0.19	0.48	0.09
<i>Forma de Pagamento</i>				
Quitação à vista	-0.07	-0.09	<b>-0.63</b>	-0.53
Quitação de 2 a 5 parcelas	0.05	0.16	0.02	<b>0.72</b>
Quitação em 6 parcelas ou mais	0.04	-0.05	<b>0.80</b>	-0.07
<i>Indenização contratada</i>	-0.11	-0.40	<b>0.72</b>	-0.19
<i>Franquia Obrigatória</i>	0.32	-0.11	0.30	0.56
<i>Gênero Masculino</i>	-0.33	0.20	0.13	<b>0.67</b>
<i>Tipo de Renovação</i>				
Renovação de congênere	-0.22	<b>-0.75</b>	0.13	0.21
Renovação própria	-0.02	<b>0.94</b>	-0.04	0.01
Seguro novo	0.28	<b>-0.74</b>	-0.08	-0.26
<i>Estado Civil</i>				
Solteiros	<b>0.97</b>	-0.05	0.10	-0.02
Não solteiros	<b>-0.97</b>	0.05	-0.10	0.02
<i>Risco</i>				
Alto risco	0.21	<b>-0.74</b>	-0.19	-0.40
Baixo risco	-0.21	<b>0.74</b>	0.19	0.40
<i>Importância Segurada</i>	0.04	0.11	<b>-0.46</b>	0.64
<i>Idade do veículo</i>	-0.15	0.35	<b>0.82</b>	0.20
<i>Proporção da Variância Explicada</i>				
Individual	0.28	0.19	0.16	0.09
Acumulada	0.28	0.46	0.63	<b>0.72</b>



A tabela acima mostra os carregamentos após a rotação, bem como a variância explicada por cada fator, e a proporção acumulada de explicação da variância. É possível observar, portanto, que os fatores considerados são capazes de explicar 72% da variância total.

Os maiores carregamentos sobre o primeiro fator foram obtidos pelos atributos que se referem à faixa etária e ao estado civil dos condutores. É importante notar que altos escores neste fator representam os condutores solteiros jovens, e que escores de menor valor representam os condutores idosos casados. Este fator que, devido às suas características, será denominado fator de estrutura familiar, foi aquele que individualmente proporcionou maior explicação da variância, com 28% de explicação.

Os atributos que apresentaram maior carga sobre o segundo fator representam o tipo de renovação e o risco, medido através da classe de bônus do segurado. É interessante observar que este fator assume maior escore em casos onde as renovações são da própria companhia e a classe de bônus é elevada. Este fato pode ser interpretado como a fidelidade dos segurados. O comportamento da fidelidade pode estar sendo influenciado pelos descontos referentes a renovações da própria companhia, conjugados com a política de bonificação adotada. Outro fato interessante é que os menores escores representam as renovações de congêneres e seguros novos, com maior risco. Supõe-se que este fato seja reflexo da estratégia da companhia diante das demais seguradoras. Provavelmente os tipos de veículos que apresentaram menor escore neste fator são aqueles em que o seguro da companhia assume preços compensadores diante do mercado. Através da distribuição de carregamentos observada neste fator, decidiu-se por denominá-lo fator de estratégia de mercado e risco.

O terceiro fator pode ser entendido como a condição financeira dos segurados. Isto provém do fato de que escores elevados representariam contratos convencionais, para veículos de maior idade e de menor valor, e ainda quitados através de um número maior de prestações. Ou seja, os maiores escores representam os segurados com menor poder aquisitivo, vice-versa.

O último fator a ser selecionado apresentou como relevantes os atributos referentes ao sexo e à quitação em poucas parcelas. Este fator, dentre os demais, foi o que explicou a menor proporção da variância, com apenas 9% de explicação.

Como a forma de pagamento de maior carga foi dentre as existentes, a mais comum, este fator foi interpretado como representante do sexo do condutor.

Diante do exposto, conclui-se que os atributos proporcionais utilizados para representar os tipos de veículos, podem ser expressos em termos de 4 fatores, que explicam 72% da variância total. São eles: o fator estrutura familiar, o fator estratégia de mercado, o fator condições econômicas e o fator sexo do condutor.

Construídos os fatores e identificado seu comportamento, os mesmos foram utilizados como variáveis, no agrupamento dos tipos de automóveis.

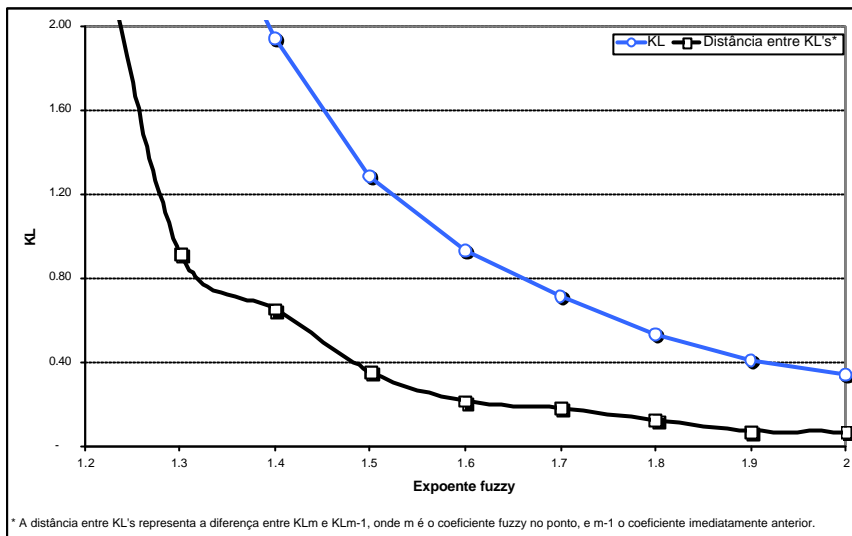
Conforme apresentado na seção 3.1, o método de agrupamento utilizado é conhecido como *fuzzy c-means* (FCM).

Para iniciar o FCM é necessário definir o número de agrupamentos, o critério de parada, o coeficiente *fuzzy* e a matriz contendo os graus de pertinência iniciais.

Sabendo que os agrupamentos definidos neste instante não seriam mantidos, já que, a posteriori, espera-se que este número seja reduzido, decidiu-se obter do FCM uma estrutura contendo 20 agrupamentos de veículos. Além disso, foi utilizado como critério de parada o valor  $10^{-4}$ . A matriz inicial de graus de pertinência foi selecionada de forma aleatória.

Para a definição do coeficiente *fuzzy* foi utilizada a média da entropia de Kullback-Liebler. Conforme definida na seção 3.1, esta medida é aplicada quando é necessário identificar se duas distribuições de probabilidade divergem. Portanto, buscou-se analisar graficamente o comportamento desta medida, diante da utilização de coeficientes *fuzzy*.

Figura 13: Medida de Kullback-Liebler, diante dos graus de pertinência.



De acordo com o gráfico, o primeiro momento de estabilidade da medida foi alcançado entre 1.7 e 1.6. Por este motivo, como a intenção era de obter o agrupamento mais *fuzzy* possível, porém com capacidade de discriminar a pertinência de cada indivíduo a cada agrupamento, decidiu-se pelo coeficiente *fuzzy* de valor 1.7.

Definidos os parâmetros, o FCM determinou 20 agrupamentos, de acordo com os fatores pré-estabelecidos. A interpretação dos grupos formados, bem como o valor dos fatores que determinam o centro dos grupos, encontram-se na tabela 12 abaixo.

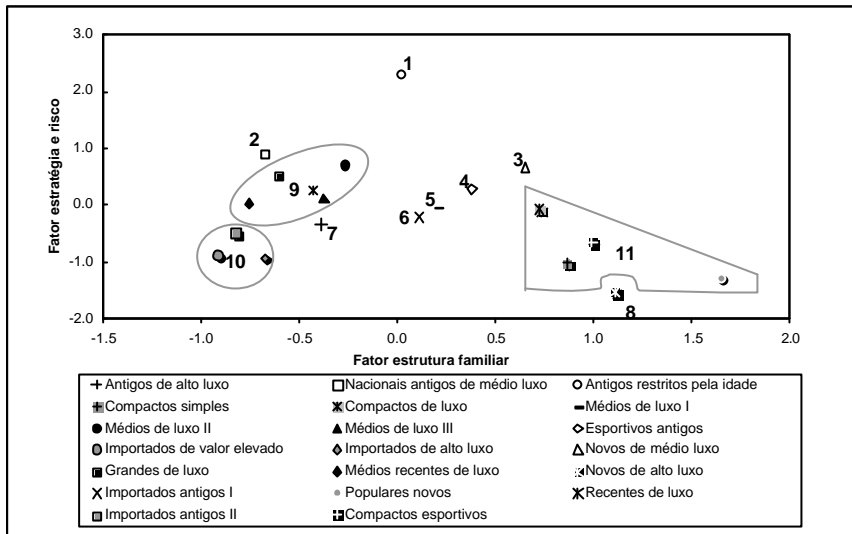
Tabela 8: Interpretação dos grupos de veículos formados na iniciação e valor assumido pelos centróides.

	Estrutura Familiar	Estratégia e Risco	Situação Financeira	Sexo do Condutor
Compactos esportivos	0.99	-0.64	1.37	0.01
Antigos de alto luxo	-0.39	-0.33	0.87	0.13
Nacionais antigos de médio luxo	-0.68	0.90	1.02	-0.14
Antigos restritos pela idade	0.02	2.29	1.09	-0.32
Compactos simples	0.86	-1.02	0.34	-0.55
Compactos de luxo	0.72	-0.07	-0.19	-0.64
Médios de luxo I	0.21	-0.05	-1.01	-0.71
Médios de luxo II	-0.27	0.70	0.07	-0.28
Médios de luxo III	-0.38	0.12	-0.93	-0.03
Esportivos antigos	0.38	0.28	1.04	-0.29
Importados de valor elevado	-0.92	-0.89	-0.66	1.88
Importados de alto luxo	-0.67	-0.93	0.04	1.14
Novos de médio luxo	0.65	0.67	-0.39	-0.55
Grandes de luxo	-0.61	0.52	-0.24	-0.03
Médios recentes de luxo	-0.76	0.03	-1.42	-0.52
Novos de alto luxo	1.10	-1.52	-0.74	0.81
Importados antigos I	0.11	-0.21	-0.03	0.20
Populares novos	1.65	-1.29	-0.15	-0.66
Recentes de luxo	-0.43	0.25	-2.45	-2.64
Importados antigos II	-0.83	-0.48	0.52	0.75

A tabela acima mostra que, embora grande parte dos grupos apresente características próprias, existem casos, como o caso dos grupos 7, 8 e 9, em que as semelhanças inviabilizaram interpretações claras, e, provavelmente, nestes casos a divisão não se justificaria.

Esclarecendo a relação entre os grupos e possibilitando a interpretação da relação entre os fatores e os agrupamentos formados, o gráfico abaixo mostra o posicionamento dos grupos considerando os fatores estrutura familiar e estratégia e risco.

Figura 14: Centros dos agrupamentos segundo os fatores estrutura familiar e estratégia e risco.



Na figura acima, os objetos determinam os grupos, as linhas claras representam que há uma associação por semelhança entre grupos vizinhos, e a numeração identifica cada associação o grupo isolado.

Considerando os dois fatores de maior explicação, a figura acima mostra que os grupos seguem um determinado padrão lógico.

Diante do valor assumido pelos fatores, é possível notar que veículos antigos restritos (1) só sejam aceitos mediante um bom conhecimento do risco. Porém, embora exista a suspeita de que condutores com maiores responsabilidades familiares apresentem menor risco, este fato não influencia na aceitação de tais veículos. Outra observação relevante seria notar que estes veículos apresentam comportamento distinto dos demais, fato que é representado pela distância entre este grupo, e os grupos restantes.

Verifica-se que veículos antigos de médio luxo (2) e novos de médio luxo (3), embora apresentem riscos e estratégias semelhantes por parte da companhia, demonstram características particulares em relação à estrutura familiar. Enquanto pessoas mais velhas e de maior responsabilidade familiar preferem veículos antigos, veículos novos têm a preferência dos jovens. Este fato pode ser entendido pelo próprio design que diferencia estes veículos.

Veículos esportivos antigos (4) são conhecidos pela baixa liquidez de mercado, e por representarem um pseudo-status econômico-social. A baixa

liquidez faz com que estes veículos assumam baixo valor de mercado. Este fato, associado ao luxo e à potência, torna tais veículos atraentes ao consumidor mais jovem. Estes veículos também podem ser enquadrados no perfil de restrição, e por este motivo só costumam ser aceitos mediante baixo risco e conhecimento prévio das características dos condutores.

A proximidade entre veículos importados (6) antigos e veículos médios de luxo (5) pode ser resultante do próprio valor de mercado destes veículos. Nota-se ainda que ambos apresentam características intermediárias em relação ao risco e à estrutura familiar.

Interessante notar que veículos antigos de alto luxo (7) e veículos importados (6) apresentam perfil semelhante em relação ao risco, porém características distintas entre os condutores.

Os veículos importados (10) se assemelham em muito aos veículos antigos de alto luxo. Este fato pode ser entendido pelas semelhanças de luxo e design entre os mesmos. Estes dois grupos compõem a massa de veículos conhecida no mercado de seguros pelo seu alto risco. Neste caso, alto risco não representa apenas uma maior probabilidade de sinistros, mas também sinistros que envolvem altas indenizações, e conseqüentemente, despesas elevadas para as seguradoras. Por este motivo, grande parte destes veículos remanesce de outras seguradoras, conforme representado pelos valores do fator de estratégia e risco.

A proximidade entre os fatores mostra ainda a relação pouco surpreendente entre veículos novos e recentes de luxo (9).

Normalmente é interessante para as seguradoras, mesmo tendo que abrir mão do conhecimento prévio do risco, e do risco a priori de condutores jovens e solteiros, admitir veículos compactos e populares (11). Este interesse é resultante da representatividade destes veículos diante da frota circulante brasileira.

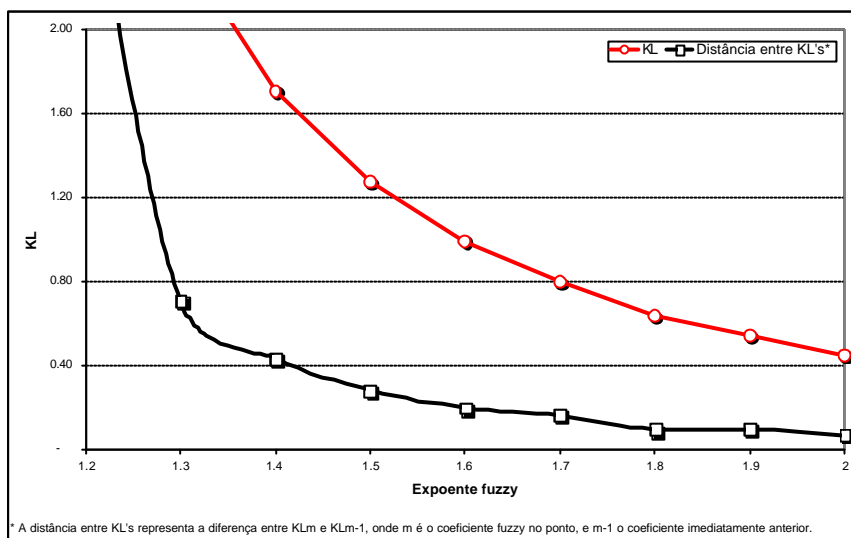
Finalmente, os veículos novos de alto luxo (8), acabam por representar o perfil que, embora aceito pelas seguradoras, é o mais indesejado. Veículos normalmente com perfil esportivo, de alta potência e luxo, e que são conduzidos, na maioria dos casos, por jovens de classe alta, formam o perfil de altíssima probabilidade de sinistros e indenizações de valores exorbitantes.

Seguindo para o agrupamento inicial das regiões de tarifação, a mesma técnica de agrupamento de dados aplicada no caso dos tipos de veículos, foi utilizada.

Para a formação dos agrupamentos iniciais de regiões, decidiu-se pela utilização de variáveis que representassem a frequência dos três tipos de sinistros considerada por este estudo. Através destas variáveis, busca-se diferenciar as regiões de tarifação de acordo com a maior incidência, diante das unidades expostas, de sinistros de perda parcial, perda total e roubo. Cabe ressaltar que as variáveis foram padronizadas na tentativa de evitar a influência de uma ou outra variável isoladamente no agrupamento.

Dando início ao agrupamento, e adotando a mesma estratégia de definição do coeficiente *fuzzy* adotada no caso dos tipos de veículos, foi utilizada a média da entropia de Kullback-Liebler.

Figura 15: Medida de Kullback-Liebler, segundo os graus de pertinência utilizados.



A figura acima apresenta aproximadamente o mesmo comportamento apresentado pelo gráfico 12. Por este motivo, decidiu-se que o coeficiente *fuzzy* a ser utilizado na agregação e classificação de regiões, assumiria o mesmo valor daquela ocasião. Além deste, com exceção do número de agrupamentos, também foram considerados os mesmos parâmetros de iniciação do FCM.

No caso das regiões foi determinada a construção de 7 agrupamentos. A interpretação dos agrupamentos formados, bem como os valores que determinam o centro dos grupos, encontram-se na tabela 13 abaixo.

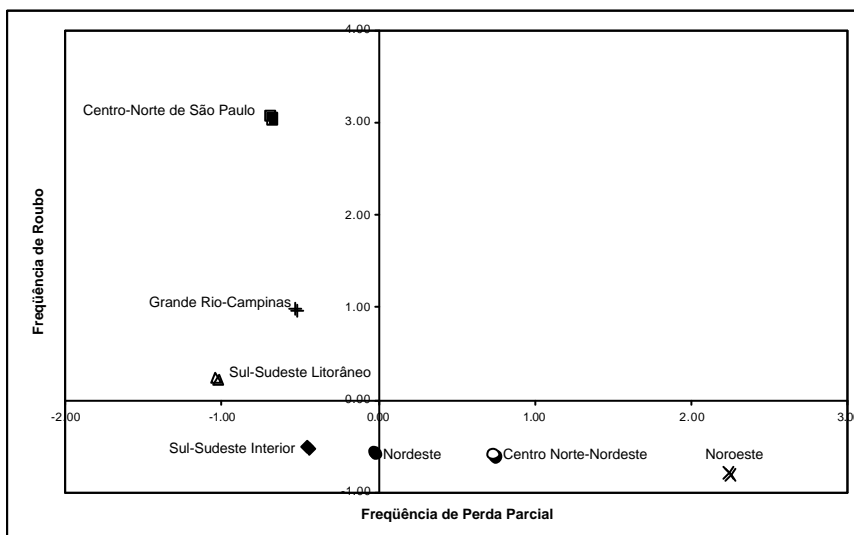
Tabela 9: Interpretação dos grupos de regiões formados na iniciação e valor dos centróides.

	Perda Parcial	Perda Total	Roubo
Grande Rio-Campinas	-0.53	-0.22	0.99
Sul-Sudeste Litorâneo	-1.04	-0.54	0.24
Centro-Norte Nordeste	0.73	-0.19	-0.59
Centro-Norte de São Paulo	-0.69	-0.73	3.07
Sul-Sudeste Interior	-0.45	-0.27	-0.50
Nordeste Indefinido	-0.03	1.53	-0.56
Noroeste	2.23	1.15	-0.79

Na intenção de compreender as diferenças entre os agrupamentos formados, foi construído um gráfico representando os centros dos grupos.

Embora tenham sido utilizadas no agrupamento, tanto a frequência de perdas parciais quanto a frequência de perdas totais, verificou-se que estas apresentam-se correlacionadas, cujo coeficiente de correlação atingiu o valor de 0.62, aproximadamente. Este fato é compreensível, já que ambos os danos costumam ser causados por colisões, portanto causados por fatores semelhantes, diferindo apenas no grau do dano causado. Por este motivo a frequência de perda total foi omitida na análise gráfica, cujos eixos representam apenas as frequências padronizadas de perda parcial e roubo.

Figura 16: Centros dos agrupamentos segundo a frequência de perdas parciais e roubo.





Analisando o gráfico acima, verifica-se que as regiões, de uma maneira geral, se diferenciam bem considerando quando consideradas apenas as frequências de perda parcial e roubo.

De acordo com o gráfico, os agrupamentos definidos como Centro-Norte de São Paulo, Grande Rio-Campinas e Sul-Sudeste Litorâneo apresentam frequências de sinistros de perda parcial muito semelhantes, diferenciando-se apenas pela frequência de roubo. Acredita-se que os altos índices de roubo apresentados por estas regiões tenham relação com a densidade populacional e a desigualdade social.

Comportamento inverso pode ser percebido quando são analisadas as Regiões Sul-Sudeste Interior, Nordeste, Centro Norte-Nordeste, e Noroeste. As mesmas apresentam aproximadamente a mesma frequência de roubo, porém se distinguem pelas variações na frequência de perda parcial. Acredita-se que a frequência de perdas parciais esteja relacionada com as condições de conservação das estradas e vias públicas destas regiões.

Finalizada a exposição dos agrupamentos formados tanto no caso dos tipos de veículos, quanto no caso das regiões de tarifação, considera-se relevante expor a definição da lógica de agrupamento das idades dos veículos que será aplicada, na medida em que se torne necessário.

O agrupamento das idades segue uma estratégia diferente da utilizada, tanto no caso dos tipos de veículos, quanto no caso das regiões de tarifação.

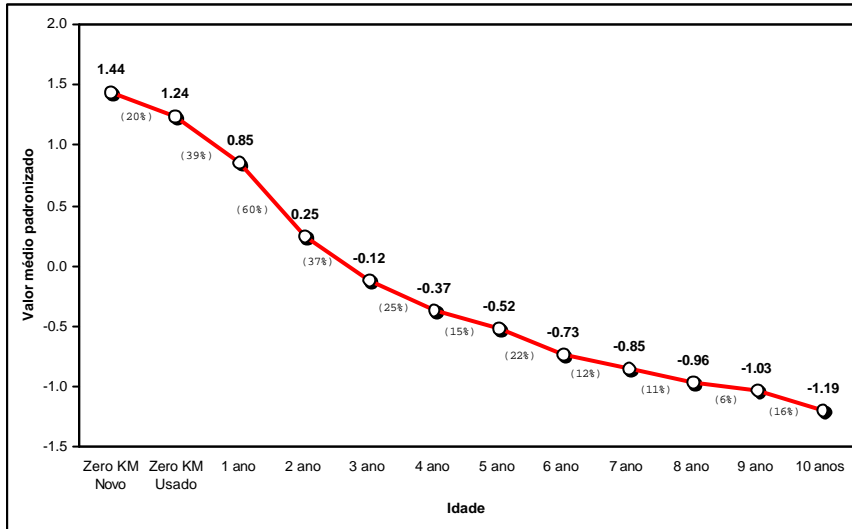
Neste caso, a intenção é medir a perda de valor atrelada ao tempo de vida dos veículos, e formar agrupamentos com base nas diferenças relativas entre os valores médios dos mesmos. Esta proposta surge da necessidade de agrupar as idades de maneira lógica, ou seja, seguindo um raciocínio contínuo. Seria indesejável que, por exemplo, veículos de 10 anos ou mais acabassem por se agrupar com veículos zero quilômetro. O desejável seria que os mesmos se agrupassem com veículos de 9 anos.

Inicialmente, padronizou-se o valor de cada tipo de veículo, em cada ano de fabricação, utilizando seu valor médio de mercado e a dispersão, medida pelo desvio padrão, entre as idades. Em seguida, foi calculada a média dos valores padronizados de todos os veículos, em cada uma das idades. Pretende-se, com a padronização que, veículos lançados recentemente, bem como veículos de maior valor de mercado, tenham a mesma influência sobre a média por idade. A

distância entre valores será utilizada para nortear o agrupamento das idades, pois mostra a proximidade entre de valor entre as mesmas.

As figuras abaixo mostram a curva de perda de valor e o comportamento da distância entre idades, respectivamente.

Figura 17: Tendência média da perda de valor de mercado dos veículos.



A figura 17 mostra, acima de cada ponto, o valor médio de mercado padronizado. Abaixo, entre os pontos, mostra a diferença relativa dos valores entre idades.

Observando a figura, nota-se que a perda de valor apresenta tendência decrescente, conforme esperado.

É possível notar ainda, que com exceção dos veículos novos usados, que representam um nível entre zero quilômetro e veículos com 1 ano de fabricação, todas as demais idades diferem entre si, de um mesmo período de tempo. Logo, em um primeiro momento, se esperaria que a perda de valor fosse aproximadamente constante. Ou seja, que a distância relativa tivesse o mesmo valor, no decorrer do tempo. Porém, verifica-se que as diferenças oscilam, decrescendo, ano após ano. Este comportamento pode ser entendido como o comportamento da oferta e da demanda do mercado por veículos mais antigos. Veículos novos usados, conforme mostra a figura 17, perdem, em média, 20% do seu valor de mercado, a partir do momento em que passam a circular. Após o 1º

ano de utilização, a tendência é de que, ano após ano, o veículo desvalorize menos em relação aos anos anteriores.

Conforme observado, seguindo a linha do tempo de utilização e de acordo com o valor de mercado dos veículos, a lógica mostra que aqueles com maior tempo de utilização deveriam se agrupar com os mais recentes. Porém, este comportamento não pode ser generalizado. Um exemplo seriam os veículos com 2 anos de utilização. Neste caso, ao invés de agrupá-los com veículos mais recentes, no caso, veículos com 1 ano de utilização, a figura mostra que veículos com 2 anos deveriam ser agregados aos veículos com 3 anos, já que entre 2 e 3 anos a diferença relativa é de 37%, e entre 1 e 2 anos a diferença é de 60%.

A lógica de agrupamento proposta para as idades dos veículos, segue a estrutura exemplificada acima, e será aplicada em momento oportuno.

#### **4.3.2. Evolução e características do algoritmo**

Esta seção se propõe a apresentar as características e a evolução do algoritmo de classificação proposto.

Antes de iniciar o algoritmo, a base de dados original contava com 34.475 classes de risco, onde 64.2% delas não apresentavam sinistros, e aproximadamente 28% apresentavam exposição inferior à unidade. Conforme foi mostrado na seção 2.2.1, suspeita-se que o percentual de classes sem sinistros e o percentual de classes com exposição inferior à unidade sejam conseqüências da classificação individualizada. Acredita-se que com o número reduzido de informações de sinistros por classe, a confiabilidade das estimativas do risco ficaria comprometida. Outro ponto prejudicial à estimação seria o número de *outliers* gerados quando as classes de risco apresentam informação de sinistros e exposições baixas. Tomando por base estas considerações, o método de classificação proposto visa reduzir, quando não extinguir, estes problemas.

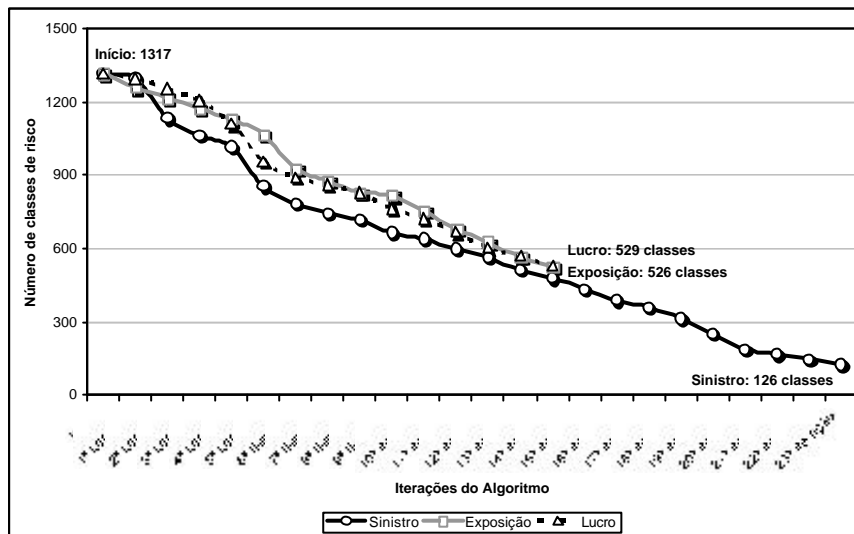
Conforme sugeridos na seção 3.1, três critérios foram aplicados no intuito de diluir a classificação do risco. O primeiro critério estabelece o encerramento do algoritmo caso nenhuma classe apresente exposição inferior à unidade. Como forma de simplificação, o algoritmo regulado por este critério será denominado “*algoritmo da exposição*”.

O segundo critério aplicado determina o encerramento do algoritmo caso o MPE esteja contido no intervalo  $[1\%,1\%]$ . Como esta medida de erro pode ser entendida como o excedente de lucro na estimação do risco, o algoritmo controlado por este critério será denominado “*algoritmo do lucro*”.

Finalmente, o terceiro critério visa a extinção de classes sem sinistros, e será denominado “*algoritmo de sinistro*”<sup>33</sup>.

Os algoritmos foram iniciados com 1.317 classes de risco, formadas através da conjugação dos 20 agrupamentos de veículos com 7 grupos de regiões e 12 idades de veículos<sup>34</sup>.

Figura 18: Evolução do número de classes de risco.



A figura acima mostra a evolução do número de classes de risco formadas desde a iniciação dos algoritmos até serem finalizados. Nota-se que os algoritmos da exposição e do lucro finalizaram na 14ª iteração, e que o algoritmo de sinistro finalizou na 23ª iteração. Este fato corrobora a expressividade do número de classes sem sinistros.

De acordo com a figura 18, verifica-se que, ao final dos algoritmos, a classificação de risco, quando comparada à classificação inicial, sofreu uma redução no número de classes superior a 59%. Nota-se ainda que até a 14ª iteração

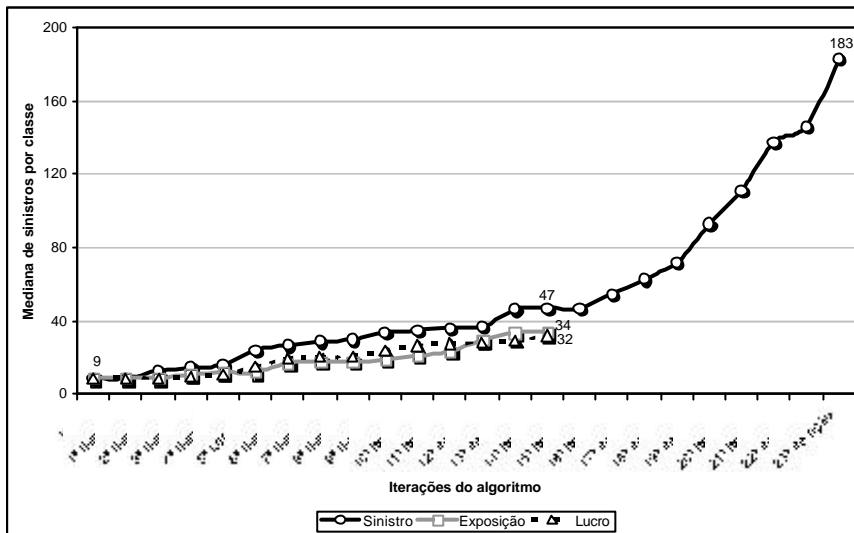
<sup>33</sup> Será tomada liberdade em utilizar os termos critério e algoritmo, acreditando que esta decisão não dificultará, ao leitor, a compreensão das análises. Estes dois termos também serão utilizados nos demais capítulos.

<sup>34</sup> Neste caso, poderiam ser construídas 1.680 classes de risco.

o número de classes obtido pelos algoritmos foi semelhante. Além disso, nota-se que o menor número de classes final foi obtido pelo algoritmo de sinistro, que apresentou 126 classes de risco.

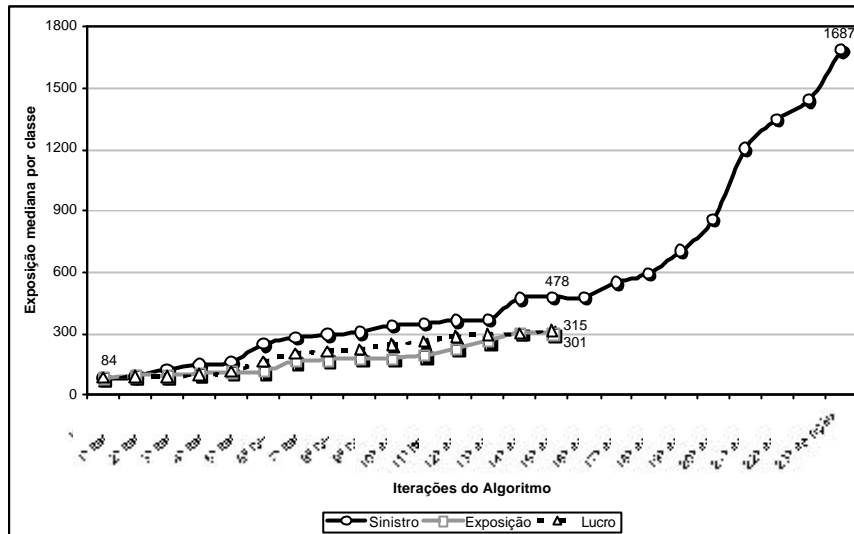
Como o objetivo principal dos algoritmos é a obtenção de um volume maior de informações de sinistros e exposição por classe, decidiu-se investigar a evolução dos valores medianos destas variáveis<sup>35</sup>. Os resultados encontram-se representados nas figuras 19 e 20, abaixo.

Figura 19: Evolução do número mediano de sinistros por classe.



<sup>35</sup> Optou-se por utilizar a mediana, ao invés da média, devido à suposição de as distribuições em análise sejam assimétricas.

Figura 20: Evolução da exposição mediana por classe.



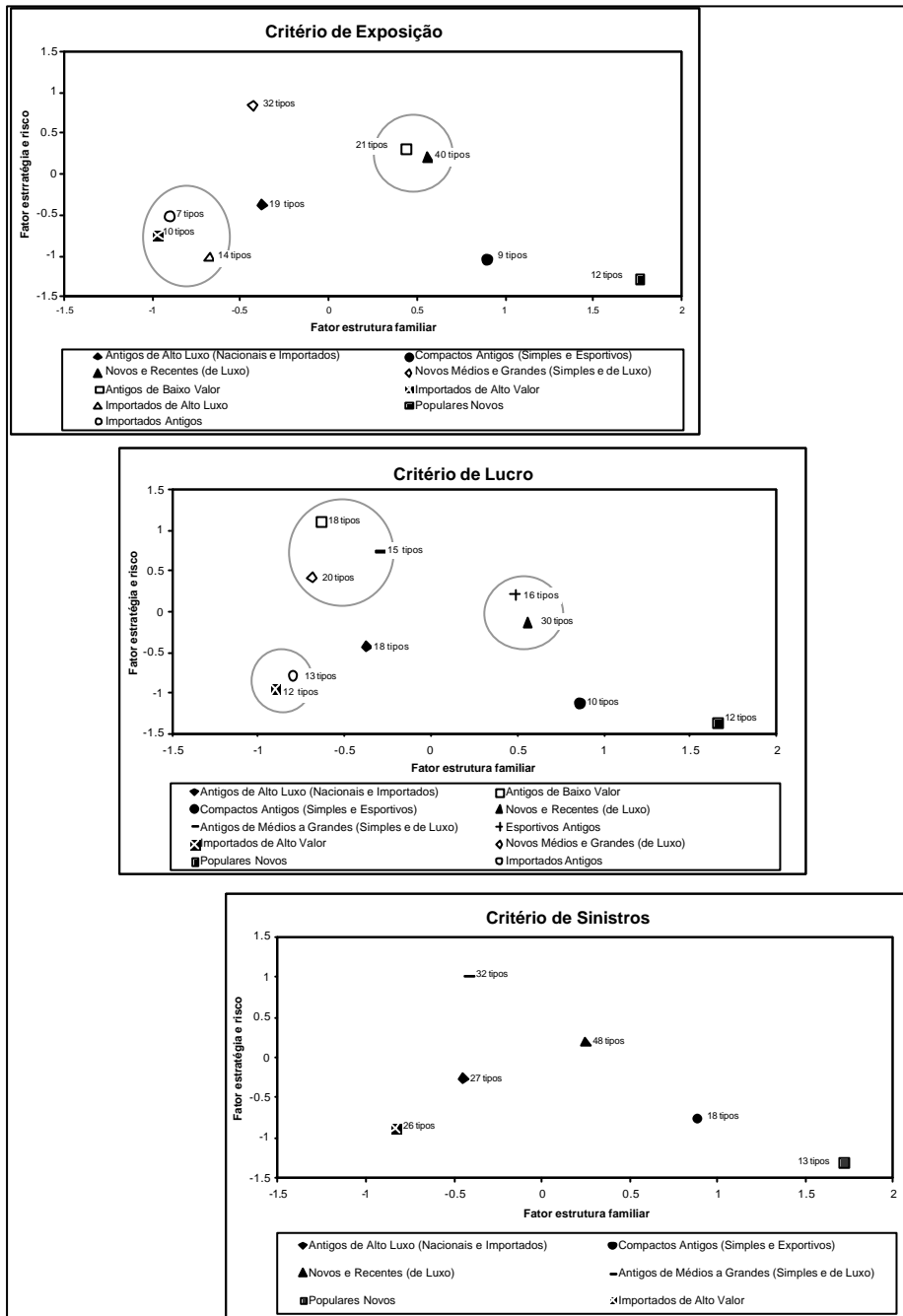
Nas figuras acima, os pontos em destaque representam o número mediano de sinistros e a exposição mediana obtidos ao final dos algoritmos.

Observando os gráficos, nota-se que na 14ª iteração, os algoritmos do lucro e da exposição não apresentaram diferenças de agregação relevantes. Porém, nesta mesma iteração, o algoritmo de sinistro, com 478 itens expostos e 47 sinistros, apresentou, comparado aos demais algoritmos, 51% a mais de exposição e 38% a mais de sinistros. Diante deste fato, e considerando o objetivo do método proposto, acredita-se que o algoritmo de sinistro seja mais eficiente que aos demais.

Após analisar a evolução dos algoritmos, seria relevante avaliar os agrupamentos finais obtidos.

Visando compreender o comportamento dos algoritmos ao agrupar tipos de veículos, a figura abaixo mostra o posicionamento dos centróides que representam os grupos formados.

Figura 21: Agrupamentos de tipos de veículos diante dos critérios testados.



A figura acima representa o agrupamento final obtido pelos algoritmos. Os marcadores representam os centróides dos grupos, e as circunferências de cor clara, presentes apenas nos gráficos de exposição e lucro, representam casos onde se verifica maior proximidade entre grupos.

Embora a proximidade entre os centróides, detectada na finalização dos algoritmos de exposição e lucro, sugira que os grupos formados se agregariam e, conseqüentemente, resultariam nos agrupamentos obtidos a partir do algoritmo de sinistro, isto não se confirma. Este fato pode ser verificado se comparados os números de tipos de veículos que compõe os agrupamentos.

É importante ressaltar ainda que, embora os agrupamentos representados nos gráficos, em muitos casos, apresentem a mesma denominação, os tipos de veículos que formaram os grupos não são os mesmos<sup>36</sup>.

Comparando a agregação promovida pelo algoritmo de sinistro às demais, nota-se que este facilita a compreensão e a interpretação dos grupos. Interpretando os agrupamentos promovidos por este algoritmo, torna-se claro que veículos Compactos Antigos e Populares Novos são, em geral, conduzidos por pessoas jovens, solteiras e de maior risco, mas que por motivos estratégicos<sup>37</sup>, acabam por migrar de outras seguradoras. Verifica-se ainda que os veículos Importados de Alto Valor são representados por condutores mais velhos, de maior risco e que, na maioria dos casos, também migram de outras companhias.

É fato que os veículos Antigos de Médios a Grandes são restritos pela maioria das seguradoras brasileiras. Acredita-se que somente diante do bom conhecimento do risco as companhias acabem por aceitá-los. Suspeita-se que, por este motivo, tais veículos sejam, em grande parte, resultado de renovações de contrato internas.

Concluindo a análise dos agrupamentos, verifica-se que veículos Novos e Recentes e Antigos de Alto Luxo acabam por serem indiferentes aos fatores considerados no gráfico. Suspeita-se que a distinção entre estes dois grupos seja representada principalmente pelo fator que indica a situação financeira dos segurados.

Para descrever o comportamento dos algoritmos ao agrupar regiões, a figura abaixo segue o mesmo padrão utilizado na análise dos grupos de tipos de veículos.

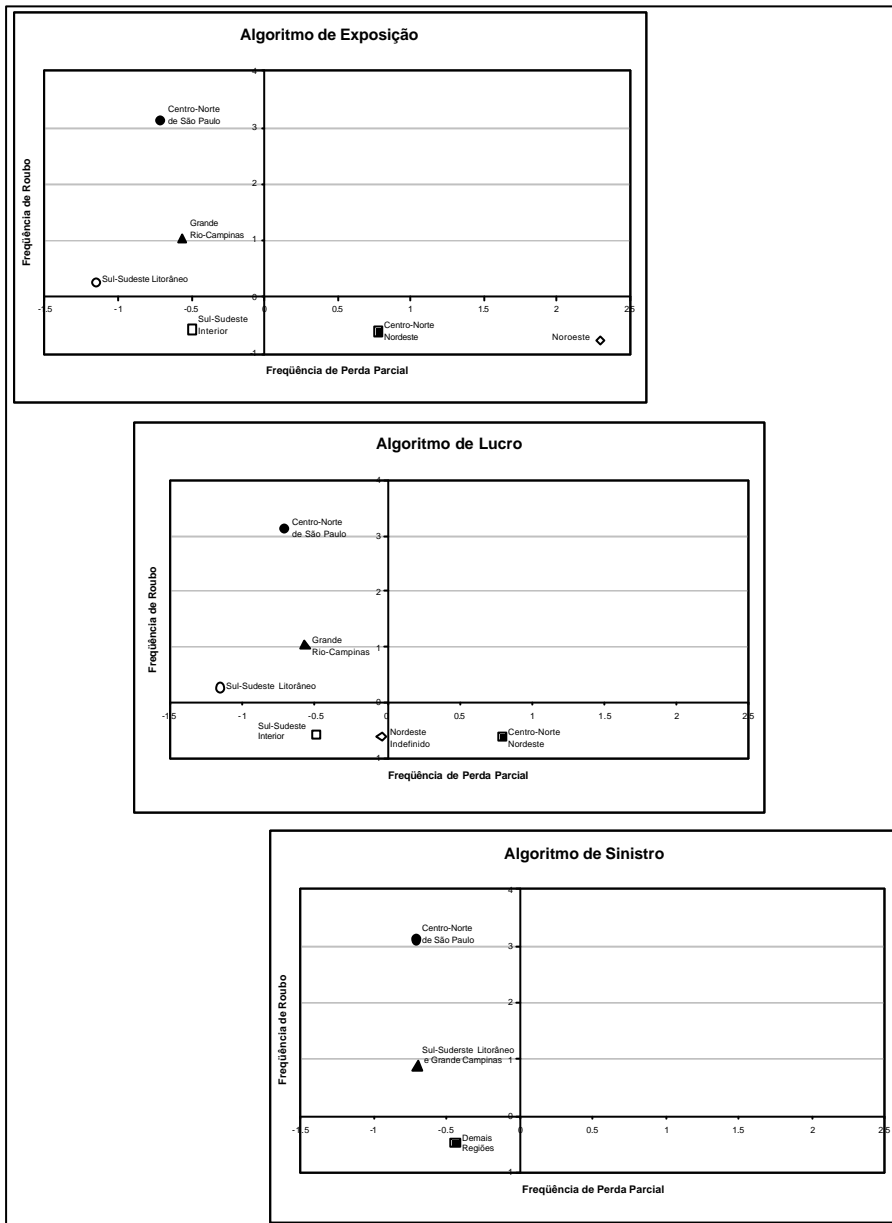
---

<sup>36</sup> A denominação dos grupos foi definida considerando as características mais freqüentes.

<sup>37</sup> Os veículos Compactos Antigos, principalmente esportivos, representam um risco indesejável. Suspeita-se que a migração destes veículos seja conseqüência da política de restrição ao risco adotada por outras seguradoras. Já os veículos Populares Novos, por representarem grande parte da frota segurada brasileira, estão sempre envolvidos em estratégias de preços.



Figura 22: Agrupamentos finais de regiões de tarifação.



Diante dos resultados representados na figura acima, nota-se que o algoritmo da exposição e o algoritmo do lucro formam agrupamentos semelhantes. Ambos criam três níveis de distinção em relação ao roubo, e quatro níveis em relação às perdas parciais. O algoritmo de sinistros considera a mesma distinção em relação ao roubo, porém não se mostra sensível às diferenças em relação às perdas parciais. Então surge a hipótese de que o algoritmo de sinistro seja menos eficiente que os demais em agrupar regiões.

Tabela 10: Características dos agrupamentos.

Algoritmo e centróides	Itens expostos	Frequência de sinistros por unidade exposta (%)		
		Perda Parcial	Perda Total	Roubo
<b>Exposição</b>				
Centro-Norte de São Paulo	81008	7.90	0.53	2.13
Centro-Norte Nordeste	67929	10.82	0.86	0.38
Grande Rio-Campinas	146450	7.85	0.89	1.51
Noroeste	10235	13.20	1.31	0.34
Sul-Sudeste Interior	88183	7.91	0.77	0.38
Sul-Sudeste Litorâneo	34202	6.70	0.73	0.80
Média	71335	9.06	0.85	0.92
<b>Lucro</b>				
Centro-Norte de São Paulo	81008	7.90	0.53	2.13
Centro-Norte Nordeste	76754	11.14	0.90	0.38
Grande Rio-Campinas	146450	7.85	0.89	1.51
Nordeste Indefinido	5110	9.20	1.57	0.25
Sul-Sudeste Interior	84483	7.87	0.74	0.38
Sul-Sudeste Litorâneo	34202	6.70	0.73	0.80
Média	71335	8.44	0.89	0.91
<b>Sinistro</b>				
Centro-Norte de São Paulo	81008	7.90	0.53	2.13
Demais Regiões	160217	9.37	0.86	0.36
Sul Sudeste Litorâneo Grande Rio-Campinas	186783	7.73	0.84	1.35
Média	124836	8.36	0.78	1.19

Na tabela acima, encontram-se representadas as características dos grupos no que se refere à exposição e às frequências de perda parcial, perda total e roubo. Os valores destacados apontam às semelhanças entre grupos, quando comparados diante de uma mesma característica.

Ao observar os valores destacados na tabela 10, verifica-se que o algoritmo de sinistro é o que melhor distingue as regiões, além de promover maior equidade no que se refere à exposição.

Para analisar as características do agrupamento de idades, serão consideradas as informações representadas na tabela 11, abaixo.

Tabela 11: Agrupamentos de idades formados segundo os critérios testados.

Idade dos veículos	Início dos Algoritmos	Algoritmo da Exposição	Algoritmo do Lucro	Algoritmo de Sinistro
Zero KM Novos Usado	Zero KM Novos Usado	Zero KM ou Novos Usados	Zero Km a 1 Ano	Zero Km a 1 Ano
1 Ano	1 Ano	1 Ano		
2 Anos	2 Anos	2 Anos	2 Anos	2 Anos
3 Anos	3 Anos	3 Anos	3 Anos	3 Anos
4 Anos	4 Anos	4 Anos	4 Anos	4 Anos
5 Anos	5 Anos	5 Anos	5 Anos	5 Anos
6 Anos	6 Anos	6 Anos	6 Anos	6 Anos
7 Anos	7 Anos	7 Anos	7 Anos	
8 Anos	8 Anos	8 ou 9 Anos	8 Anos	
9 Anos	9 Anos			
10 Anos ou Mais	10 Anos ou Mais	10 Anos ou Mais	9 Anos ou Mais	7 Anos ou Mais

Na tabela acima, encontram-se representadas as idades iniciais e os grupos de idades obtidos ao final dos algoritmos. As denominações destacadas representam as agregações promovidas por cada algoritmo.

Analisando os agrupamentos de idade, verifica-se que os algoritmos formam grupos semelhantes, embora seja maior a agregação diante do critério de sinistros. Verifica-se ainda, que os algoritmos tendem a agrupar as idades posicionadas nas extremidades, e manter desagregadas as idades centrais. Acredita-se que este comportamento decorra, no caso de veículos zero quilômetro, do próprio risco, e no caso de veículos antigos, da associação entre pré-seleção e taxas cobradas.