
MEGI

MESTRADO

Estatística e Gestão de Informação

***RENTABILIDADE DOS CLIENTES DE
UMA SEGURADORA AUTOMÓVEL***

Liliana Isabel Monte Rebelo

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Estatística e Gestão
de Informação

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

RENTABILIDADE DOS CLIENTES DE UMA SEGURADORA AUTOMÓVEL

por

Liliana Isabel Monte Rebelo

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Estatística e Gestão de Informação, Especialização Análise e Gestão de Risco

Orientador: Professor Doutor Pedro Corte Real

Co-orientador: Professora Doutora Gracinda Rita Guerreiro

Novembro 2014

AGRADECIMENTOS

Em primeiro lugar os meus agradecimentos são dirigidos aos orientadores desta dissertação, Professora Doutora Gracinda Rita Guerreiro e Professor Doutor Pedro Corte Real, pelas recomendações, por tudo o que me ensinaram e pelo tempo, paciência e dedicação investidos durante todo o processo de elaboração deste projecto.

Não poderia deixar de prestar os meus agradecimentos à Seguros Logo, SA, em particular ao Doutor José Pedro Inácio pela autorização de recolha e análise dos dados dos clientes da empresa e a todos os colegas de trabalho pela disponibilidade para esclarecimentos relacionados com o negócio.

Incluo também o meu obrigado aos meus pais, irmã e amigos mais próximos pelo apoio, compreensão, preocupação e motivação que me foi dada do início ao fim do projecto.

Muito obrigado a todos aqueles que, directa ou indirectamente, contribuíram para a elaboração desta dissertação.

RESUMO

O mundo está em constante evolução e as alterações daí resultantes são notórias de ano para ano. O comércio é uma das áreas onde tais mudanças são claramente visíveis.

Foram criadas novas necessidades de consumo, entre as quais a aquisição de viatura própria e, com ela, a subscrição do obrigatório seguro automóvel. Outro elemento que foi inserido no quotidiano da população portuguesa foi a aceitação e adesão às novas tecnologias, o que veio alterar a produção e os hábitos da sociedade.

O presente estudo tem como objectivo analisar a rentabilidade dos clientes de uma seguradora, cujo principal ramo de actividade é o seguro automóvel. Para que a seguradora prospere no mercado é necessário o seu reconhecimento por parte dos consumidores e, sobretudo, que o seu negócio seja rentável e bastante competitivo face ao restante mercado. Estas condições só se verificam se forem geradas mais receitas que despesas, dando ao negócio alguma estabilidade económica.

A análise em causa permitirá conhecer as características dos clientes mais e menos rentáveis para a empresa. É de extrema importância que uma seguradora conheça o perfil dos seus clientes, pois são estes que geram receita, mas são naturalmente quem também pode originar perdas elevadas, consumindo os prémios e, por vezes, as reservas de segurança.

Assim, compreender qual o potencial dos clientes gerarem mais ou menos valias é essencial para analisar possíveis falhas no desenho dos produtos, no cálculo e construção de provisões para os sinistros, na determinação dos prémios ou sistema de bonus-malus (tarifação *a posteriori*), ou no processo de subscrição. Permitindo à seguradora uma melhoria na qualidade da tomada de decisão, pois terá bases mais sólidas na análise e gestão do risco, salvaguardando a estabilidade financeira e protegendo os interesses dos seus segurados e accionistas.

Uma das formas utilizadas para estudar a rentabilidade de clientes é com recurso a técnicas estatísticas adequadas, nomeadamente recorrendo a Modelos Lineares Generalizados, que permitem analisar a relação entre variáveis e a influência das mesmas sobre outra variável. No âmbito dos seguros a utilização de MLG permitiu às seguradoras, por exemplo, perceber que características do condutor e do veículo que mais afectam os níveis de sinistralidade, tanto em número como em custos de sinistros, que são questões fundamentais no âmbito da tarifação.

PALAVRAS-CHAVE

Actividade seguradora, seguro automóvel, rentabilidade do cliente, sinistros.

ABSTRACT

The world is constantly evolving and the changes are noticeable from year to year. The trade business is one area where such changes are clearly visible.

New consumption needs were created, including the purchase of own car and, with it, the car insurance. Another element that was inserted in the daily life of the Portuguese population was the acceptance and uptake of new technology, which has greatly alter the production and habits of society.

The present study intends to analyze the profitability of the customers of an insurance company whose core business is auto insurance. To thrive in the insurance market the company needs to be recognized by consumers and, mainly, that your business is profitable and very competitive compared to the rest of the insurance market. These conditions only occur if they are generated more revenue than expenses, giving the business some economic stability.

This analysis will reveal the characteristics of the most and the least profitable customers. It is very important that the insurance company know who their customers. They are the ones that generate revenues, but, are also those who may cause heavy losses, consuming the insurance premiums and, sometimes, the safety reserves.

Thus, understand the customers' potential to generate capital gains or losses is essential to analyze possible flaws in product design, provisions for claims, the determination of premiums or bonus-malus system (charging a posteriori), or in the underwriting process. This allows the insurer to improve the quality of decision making, and it will have a more solid foundation in the analysis and risk management, safeguarding financial stability and protecting the customers and shareholders' interests.

One ways to study the profitability of customers is using appropriate statistical techniques, including using Generalized Linear Models, which allow analyze the relationship between variables and the influence thereof on the other variable. Under insurance scope, the use of GLM allowed insurance companies understand which characteristics of the driver and the vehicle that most affect accident rates, in number and cost of claims, which are key issues in the context of pricing.

KEYWORDS

Insurance, car insurance, client profitability, claims

ÍNDICE

1. Introdução	1
2. O Mercado Segurador PORTUGUÊS – Ramo Automóvel.....	2
2.1. O Mundo dos Seguros Auto em Portugal	5
2.2. Directas vs Tradicionais.....	6
2.3. Sinistros	10
2.3.1. Tipos de danos	10
2.3.2. Cenário em Portugal	11
2.4. Risco	13
2.5. Solvência	15
2.5.1. Solvência II	16
2.6. Rentabilidade	17
2.7. Tarifação.....	19
3. Objecto de Estudo.....	21
3.1. Descrição da Empresa	21
3.2. Razões pela escolha de um canal directo	21
3.3. Perfil do cliente alvo:	22
3.4. Estratégia de negócio.....	22
3.5. Oferta Logo	23
4. Modelos Lineares Generalizados	25
4.1. Modelos Lineares Clássicos.....	26
4.2. A Família Exponencial	27
4.3. Formulação dos modelos	29
4.3.1. Regressão Logística	29
4.3.2. Regressão Gama	31
4.4. Ajustamento dos modelos	31
4.4.1. Estimacção de β	32
4.4.2. Testes de hipóteses.....	32
4.5. Seleccção e avaliação dos modelos	34
4.5.1. Seleccção dos modelos.....	35
4.5.2. Avaliação dos modelos	38
5. Modelação da Rentabilidade de Clientes	43

5.1. Carteira de Clientes.....	43
5.1.1. Tratamento de Dados	44
5.1.2. Variáveis Utilizadas	44
5.2. Produção de Lucro	46
5.2.1. Produção de Lucro - Análise Estatística das variáveis	46
5.2.2. Produção de Lucro - Resultados	49
5.2.3. Produção de Lucro - Análise dos Resíduos	55
5.3. Estimação da Probabilidade de Produzir Lucros.....	57
5.4. Rentabilidade Gerada	60
5.4.1. Rentabilidade Gerada - Análise Estatística das variáveis.....	60
5.4.2. Rentabilidade Gerada - Resultados	63
5.4.3. Rentabilidade Gerada - Análise dos Resíduos	69
5.5. Estimação da Rentabilidade Gerada	71
6. Conclusões	74
7. Limitações e Recomendações para Trabalhos Futuros	75
8. Bibliografia	76
9. Anexos.....	78
9.2. Inputs e Outputs do R	91
9.2.1. Declaração de Variáveis	91
9.2.2. Análise Estatística das variáveis.....	91
9.2.3. Produção de Lucro	91
9.2.4. Rentabilidade Gerada	96
9.3. Glossário.....	100

ÍNDICE DE FIGURAS

Figura 2.1 - Número de veículos seguros em Portugal entre 1980 e 2010	5
Figura 2.2 - Quota de mercado das Seguradoras Directas em Portugal em PBEs (2007-2013)	9
Figura 2.3 - Prémios Emitidos e Custos com sinistros entre 2009 e 2012 (milhares)	11
Figura 2.4 - Evolução das Vítimas mortais (1951-2012)	12
Figura 2.5 - Taxa de sinistralidade entre 2009 e 2012.....	14
Figura 2.6 - Estrutura do Balanço	17
Figura 5.1 - Produção de Lucro - Coberturas	49
Figura 5.2 - Resíduos de Pearson Padronizados vs Alavancagem	55
Figura 5.3 - Desvios Residuais vs Valores Ajustados.....	56
Figura 5.4 - Rentabilidade média por Produto e Canal	62
Figura 5.5 - Rentabilidade Gerada	62
Figura 5.6 - Rentabilidade Gerada com barreira de outliers ligeiros	63
Figura 5.7 - Ajustamento dos dados a uma distribuição Gama.....	64
Figura 5.8 - Resíduos de Pearson Padronizados vs Alavancagem	69
Figura 5.9 - Desvios Residuais vs Valores Ajustados.....	69

ÍNDICE DE TABELAS

Tabela 2.1 - Prémios Brutos Emitidos em Portugal (milhares de euros) – 2012.....	2
Tabela 3.1 - Oferta LOGO; Produtos e Coberturas	24
Tabela 4.1- Funções de Ligação Canónica mais comuns	28
Tabela 5.1 - Variáveis Independentes.....	45
Tabela 5.2 - Produção de Lucro e Rentabilidade Gerada	46
Tabela 5.3 - Análise Preliminar - Carteira (variáveis numéricas).....	47
Tabela 5.4 - Critérios de Informação de Akaike.....	55
Tabela 5.5 - Características do <i>Cliente Padrão</i>	58
Tabela 5.6 - Probabilidade de produção de lucro – Exemplos	58
Tabela 5.7 - Rentabilidade gerada	60
Tabela 5.8 - Critério de Informação de Akaike (Rentabilidade)	68
Tabela 5.9 - Características do <i>Cliente Padrão</i> (Rentabilidade)	71
Tabela 5.10 - Rentabilidade Gerada – exemplo.....	72
Tabela 9.1 - Variáveis e respectivas categorias	81
Tabela 9.2 – Produção de Lucros - Análise Preliminar	86
Tabela 9.3 - Rentabilidade Gerada - Análise Preliminar	87
Tabela 9.4 – Produção de Lucro - Coeficientes do Modelo Ajustado.....	88
Tabela 9.5 – Rentabilidade Gerada -Coeficientes do Modelo Ajustado.....	90

LISTA DE SIGLAS E ABREVIATURAS

AIC	Akaike Information Criterion
APS	Associação Portuguesa de Seguradores
ASSISTENCIA	Assistência Em Viagem
AV	Actos de Vandalismo
CCC	Choque, Colisão ou Capotamento
DP	Danos Próprios
f.d.p.	Função densidade de probabilidade
FGA	Fundo de Garantia Automóvel
FN	Fenómenos da Natureza
FR	Furto ou Roubo
INEM	Instituto Nacional de Emergência Médica
IRE	Incêndio, Raio ou Explosão
ISP	Instituto de Seguros de Portugal
MCR	Minimum Capital Requirement
MLC	Modelos Lineares Clássicos
MLG	Modelos Lineares Generalizados
PBEs	Prémios Brutos Emitidos
POC	Protecção de Ocupantes
PTU	Privação Temporária de Uso
QIV	Quebra Isolada de Vidros
RC	Responsabilidade Civil
SCR	Solvency Capital Requirement
VS	Veículo de Substituição

1. INTRODUÇÃO

Na actividade seguradora o risco associado aos clientes tem um peso bastante significativo na solvência e rentabilidade do negócio. Como tal é importante uma medição precisa e uma gestão cuidada da carteira, tanto ao nível de volume como da relação custo benefício resultante da aquisição de novos clientes. Outro aspecto que provoca grandes oscilações nas carteiras de clientes das seguradoras é a concorrência.

Em Portugal, a constante evolução dos mercados e o avanço tecnológico permitiram o aparecimento de novas empresas no mercado segurador, aumentando assim o nível de competitividade existente até então. Algumas das novas seguradoras, as seguradoras directas, apresentam uma forma mais inovadora de comercializar os seus produtos, fazendo-o apenas por telefone ou internet, não dispendo de balcões de atendimento ou mediadores como as seguradoras tradicionais. Desta forma, o capital gasto em infraestruturas e colaboradores é substancialmente mais baixo, possibilitando a oferta de produtos semelhantes a preços mais reduzidos, tornando-se mais atractivo para um maior número de possíveis clientes.

Porém, uma carteira de grandes dimensões pode não corresponder a uma boa carteira a nível de risco. Para que uma seguradora directa se mantenha no mercado e continue a praticar preços inferiores ao sistema de venda tradicional (com balcões e/ou força de vendas específica), necessita ter uma boa carteira de clientes, tanto em quantidade como em “qualidade”. A quantidade é facilmente mensurável, porém a qualidade dos clientes torna-se um tópico bastante difícil de medir, dados os poucos requisitos necessários para a subscrição do seguro por uma via directa, sem que exista sequer um contacto presencial com o cliente. Uma das formas que pode ser utilizada para tal quantificação é a rentabilidade que se obtém com cada um dos clientes.

Em ambientes competitivos, a informação sobre a rentabilidade dos clientes pode ser bastante útil para a gestão da empresa, facilitando a tomada de decisão e servindo como base para melhorar ou construir novas metodologias de actuação no mercado, tanto ao nível dos clientes como da concorrência, que cada vez se apresenta mais activa. Não basta angariar mais clientes que a concorrência; uma boa quota de mercado pode ser importante, mas no caso de uma seguradora não é o suficiente para que esta se imponha no mercado, pois clientes que não sejam rentáveis e tenham um elevado grau de sinistralidade originam elevados custos para a seguradora, diminuindo os seus lucros e até mesmo os níveis de competitividade. Neste caso, mais importante que apenas a quantidade é, sem dúvida, aliá-la à qualidade dos clientes.

2. O MERCADO SEGURADOR PORTUGUÊS – RAMO AUTOMÓVEL

A actividade da uma seguradora tem por base o contrato de seguro, que consiste num acordo celebrado entre duas partes, através do qual a companhia de seguros (segurador) assume a cobertura de determinados riscos, comprometendo-se a cobrir as indemnizações ou a pagar o capital seguro em caso de ocorrência de sinistros, nos termos acordados. Em contrapartida, a pessoa que subscreve o contrato (tomador de seguro) fica obrigada a pagar ao segurador o prémio acordado em contrato. O acordo é estabelecido entre a seguradora e o tomador do seguro, porém o beneficiário poderá ser uma terceira pessoa (segurado), à qual estará associado o risco. O risco encontra-se coberto pelo seguro após pagamento da totalidade ou de uma fracção do prémio estabelecido, levando a falta de pagamento à anulação do contrato.

Prémios Brutos Emitidos		
Ramo Vida		6.922.395
	Subtotal Vida	6.922.395
Ramo Não Vida	Acidentes e Doença	1.260.938
	Incêndios e Outros Danos	766.784
	Automóvel	1.569.213
	Marítimo e Transportes	31.984
	Aéreo	8.245
	Mercadorias Transportadas	24.931
	Responsabilidade Civil Geral	113.854
	Diversos	206.819
	Subtotal Não Vida	3.986.769
Total		10.905.164

Tabela 2.1 - Prémios Brutos Emitidos em Portugal (milhares de euros) – 2012

Como se pode observar na Tabela 2.1 existem dois grandes grupos de seguros, os seguros do Ramo Vida que no final de 2012 representavam cerca de 63% do total de Prémios Brutos Emitidos (PBEs) em Portugal, e os seguros do Ramo Não Vida que representavam os restantes 37% dos PBEs.

Os valores provisórios apresentados pelo Instituto de Seguros de Portugal (ISP) revelam um aumento de cerca de 20% nos PBEs no ano de 2013. O Ramo Vida apresenta uma subida prevista de 33,6%, contrariando a sucessiva quebra que se tem vindo a verificar nos últimos anos. Estes valores contrastam com a quebra prevista de 3,1% no Ramo Não Vida face ao ano de 2012. Por sua vez, o seguro automóvel

apresenta uma diminuição prevista de 5,8 % de PBEs, quebra bastante semelhante à do ano anterior (5,4%).

Dentro do Ramo Não Vida verifica-se que o seguro automóvel apresenta os valores mais elevados de PBEs, representando cerca de 38% do total de PBEs Não Vida. Estes valores podem ser reflexo da obrigatoriedade do seguro automóvel, referido no artigo 4º do Decreto-Lei 291-2007, que estabelece que:

“...Toda a pessoa que possa ser civilmente responsável pela reparação de danos corporais ou materiais causados a terceiros por um veículo terrestre a motor para cuja condução seja necessário um título específico e seus reboques, com estacionamento habitual em Portugal, deve, para que esses veículos possam circular, encontrar-se coberta por um seguro que garanta tal responsabilidade, nos termos do presente decreto-lei.”.

Esta condição de obrigatoriedade garante que, em caso de sinistro, o culpado pelo mesmo seja responsabilizado pelos danos causados a terceiros, ficando o pagamento dos prejuízos a cargo da seguradora detentora do contrato de seguro do responsável pelo sinistro. As seguradoras, por lei, assumem um máximo de capital seguro de €5.000.000 por acidentes de danos corporais e €1.000.000 por acidentes de danos materiais, existindo a possibilidade de aumentar estes valores por opção do cliente. No caso de ocorrência de um sinistro em que o responsável não tenha um seguro válido é activado o Fundo de Garantia Automóvel (FGA), que garante o pagamento dos danos ao lesado para que este não fique prejudicado. *A posteriori* o responsável pelos prejuízos terá de indemnizar o FGA, com juros acrescidos.

O valor cobrado pela seguradora, para que esta assuma os custos em caso de sinistro, designado por *prémio de seguro*, é calculado de acordo com os riscos associados ao perfil do segurado e com o nível de protecção pretendido.

No caso do seguro automóvel, o prémio é calculado de acordo com as características do segurado (condutor habitual), do objecto seguro (viatura) e as coberturas que o cliente pretende subscrever. Contudo, existem ainda outros encargos incluídos no prémio, como é o caso dos custos de aquisição e gestão do contrato e ainda as cargas exigidas por lei (custo da carta verde, imposto de selo, INEM, FGA e Prevenção Rodoviária).

Os contratos de seguro automóvel são válidos pelo período de um ano, não sendo renovados automaticamente. O valor do prémio é tarifado todos os anos devido a possíveis alterações em determinadas variáveis, como a idade, tempo de carta, idade do veículo que variam de ano para ano. Aproximadamente 40 dias antes do final do

período de vigência do contrato é comunicado ao cliente a proposta de renovação, com o valor do novo prémio a pagar. Em caso de falta de pagamento, o contrato é automaticamente anulado assim que termine o prazo do contrato em vigor, deixando este cliente de pertencer à carteira.

O seguro automóvel obrigatório por lei apenas contempla a cobertura de responsabilidade civil contra terceiros que, como já referido, responsabiliza a seguradora pelo pagamento de danos materiais e/ou corporais causados a terceiros transportados ou não. Porém, os produtos de responsabilidade civil comercializados, por norma, incluem também as coberturas de Assistência em Viagem (AV) e Protecção Jurídica, as quais podem conjugadas com as coberturas de Quebra Isolada de Vidros (QIV), Protecção de Ocupantes (POC), Privação Temporária de Uso (PTU) e Veículo de Substituição (VS), por opção da companhia ou cliente. A descrição pormenorizada destas coberturas pode ser consultada na secção Anexos (9.1.1.1-Descrição de Coberturas).

Os produtos de danos próprios, vulgarmente conhecidos como seguros “contra todos os riscos”, para além das coberturas que compõem o produto de responsabilidade civil incluem outras coberturas que visam a protecção da viatura do segurado. Por norma, este tipo de coberturas tem associadas determinadas condições estabelecidas pela empresa de seguros, como é o caso da idade e o valor máximos do veículo seguro. As coberturas geralmente comercializadas são: Choque, Colisão ou Capotamento (CCC), Incêndio, Raio ou Explisão (IRE), Furto ou Roubo (FR), Actos de Vandalismo (AV) e Fenómenos da Natureza (FN). A descrição pormenorizada destas coberturas pode ser consultada na secção Anexos (9.1.1.1-Descrição de Coberturas).

Outro factor que influencia o valor do prémio é a *franquia*. Em caso de sinistro, a franquia corresponde a uma parte pré-definida das despesas que o tomador terá de suportar, não ficando a cargo da seguradora. Deste modo, a responsabilidade por parte da seguradora será menor, o que se reflecte numa diminuição do prémio de seguro. A franquia pode ser um valor fixo ou percentual sobre o valor do capital seguro ou dano. Na componente de responsabilidade civil a seguradora paga a totalidade do prejuízo ao terceiro lesado, a posteriori o valor é ressarcido pelo tomador à companhia de seguros; na componente de danos próprios a franquia é deduzida da indemnização paga ao tomador.

2.1. O MUNDO DOS SEGUROS AUTO EM PORTUGAL

Os hábitos e necessidades da actual sociedade portuguesa pouco têm em comum com os hábitos da população há 20 anos. Como reflexo da constante evolução foram criadas novas necessidades, sendo uma delas a aquisição de pelo menos uma viatura pessoal por agregado familiar. Por vezes, a aquisição de viatura própria é feita ainda antes da compra de casa/casamento o que leva à existência de uma viatura por pessoa adulta em vez de uma por família.

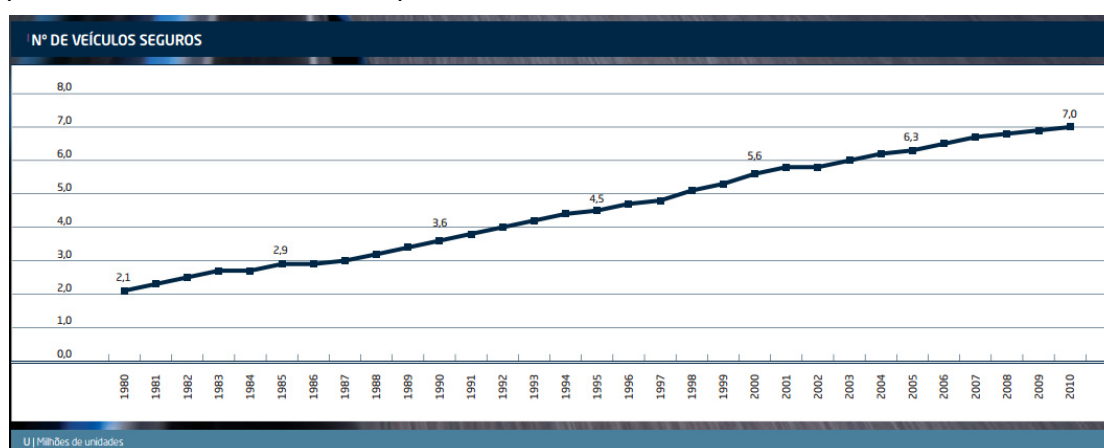


Figura 2.1 - Número de veículos seguros em Portugal entre 1980 e 2010

Dada esta evolução e uma maior oferta do mercado automóvel, o número de veículos em circulação nas estradas portuguesas aumentou consideravelmente. Como se pode observar na Figura 2.1, a circulação automóvel mais do que triplicou entre 1980 e 2010 o que, conseqüentemente, levou a um aumento da subscrição de seguro automóvel uma vez que, como referido anteriormente, se trata de um seguro obrigatório por lei. Segundo dados do ISP, o parque automóvel seguro em 2013 era composto por cerca de 6,8 milhões de veículos, verificando-se um ligeiro aumento face a 2012 (6,7 milhões), continuando no entanto abaixo dos valores de 2010 (7,0 milhões). Este decréscimo é reflexo da diminuição da venda de automóveis, mas também do aumento de veículos a circular sem seguro válido. Segundo notícia de Fonseca, Ricardo (Dezembro, 2012), publicada pela Visão, no ano de 2012 tinham sido identificados aproximadamente 20 mil casos de veículos a circular sem seguro válido, um aumento de cerca de 4% face ao ano anterior (2011).

Actualmente a procura de produtos e serviços a baixo custo é uma realidade, tendo reflexos notórios nos níveis de competitividade das empresas. Estas vêm-se obrigadas a alterar a sua oferta através de campanhas mais agressivas de modo a manter e captar mais e melhores clientes que a concorrência. Os consumidores, por sua

vez, têm à sua disposição uma oferta alargada de produtos e serviços comercializados por diversas empresas, podendo optar por aqueles que melhor se enquadram com as suas necessidades, possibilidades e preferências.

No mundo dos seguros esta realidade não é diferente. Existem inúmeras seguradoras no mercado, como tal, estas tendem a adequar a sua oferta a um maior número de consumidores.

Ao produto base comercializado, obrigatório por lei no caso do seguro automóvel, é proposto ao cliente a adição de diferentes coberturas, de acordo com as suas necessidades e ou preferências. Esta componente confere, tanto à seguradora como ao consumidor, uma maior flexibilidade a nível de risco a cobrir e consequentemente de prémios de seguro.

Com a entrada de novas seguradoras no mercado, a competitividade teve um aumento significativo que, conjugada com a facilidade e rapidez de alteração de seguradora, pode resultar numa menor retenção da carteira e flutuações indesejadas na rentabilidade da mesma.

Em ambientes competitivos a satisfação dos clientes tornou-se fundamental e, por vezes, tão ou mais importante que o preço. Dar resposta às preferências e necessidades de cada cliente é uma tarefa árdua para a generalidade das empresas. Em vários países, incluindo Portugal, existem empresas cujo foco principal é a proximidade com o cliente, enquanto outras apostam mais na simplicidade e preços baixos. Nos seguros o primeiro cenário é uma aposta das seguradoras Tradicionais, que estabelecem uma ligação mais próxima com os clientes de modo a que estes se sintam únicos. Por sua vez, as seguradoras Directas são mais generalistas, mas apresentam por regra prémios de seguro mais baratos.

2.2. DIRECTAS VS TRADICIONAIS

As seguradoras tradicionais prestam os seus serviços de atendimento com recurso a balcões e mediadores, havendo uma interação presencial com o cliente, o que possibilita um maior conhecimento sobre as suas necessidades, interesses e perfil de risco. Estes aspectos podem ser benéficos para a companhia, pois permitem uma oferta mais adequada a cada caso. No entanto, os preços praticados são, na sua maioria, mais elevados devido ao acréscimo nos custos operacionais face às seguradoras Directas.

Estas caracterizam-se por comercializarem os seus produtos através dos novos canais de distribuição - telefone e Internet - beneficiando de custos de exploração mais

reduzidos, o que lhes permite praticar, em grande parte dos casos, preços mais baixos que as seguradoras Tradicionais. Como não dispõem de mediadores como fonte de angariação de clientes, os investimentos em marketing tendem a ser mais elevados, de modo a apresentar o seu produto e respectivas vantagens face à concorrência. Esta divulgação é feita recorrendo aos media em geral, apresentando campanhas mais generalistas que promovem a marca e os produtos a um leque alargado de possíveis clientes, possibilitando uma maior exposição ao risco.

Segundo informações prestadas por Jorge Conceição Silva, administrador da Lusitania, ao Diário Económico, sobre os canais de distribuição de seguros “a diferenciação entre estes dois tipos de canal reside fundamentalmente na capacidade de aconselhamento, na complexidade das soluções e na forma e grau de envolvimento na gestão dos clientes” que, segundo opinião do mesmo, é maior numa seguradora Tradicional do que numa Directa. As companhias Tradicionais colocam à disposição dos consumidores gamas mais variadas de produtos, enquanto as Directas apresentam uma oferta mais restrita.

O grande desafio que se coloca às empresas é o de tornar o canal de distribuição uma mais-valia, de forma a apresentar os produtos e prestar serviços ajustados a cada segmento de clientes, em tempo oportuno e de forma adequada. Sublinhando esta informação, o director geral da Seguros Logo afirma que “há espaço para os diferentes tipos de distribuição continuarem a desenvolver-se, pois continuam a existir diferentes tipos de clientes, que têm necessidades diferentes” num mercado que ainda apresenta muito por explorar.

Numa fase inicial, para as seguradoras na sua generalidade os pontos mais importantes são a angariação, crescimento da quota de mercado e notoriedade da marca, só posteriormente é avaliada a rentabilidade da carteira. Dos resultados dessa análise pode ser identificada a necessidade de revisão da tarifa praticada, de modo a ajustar os prémios de seguro ao risco a cobri e assim diminuir os custos com sinistralidade.

No entanto, se a aplicação dessas novas condições se reflectir numa quebra significativa da quota de mercado, a seguradora volta a rever as condições de subscrição e tarifação de modo a entrar novamente numa fase de angariação, criando-se deste modo um ciclo de produção. Para que as alterações relativas à angariação e retenção sejam aplicadas de forma eficiente é necessário um conhecimento profundo da carteira, a fim de não por em causa a estabilidade financeira da companhia.

No momento em que a dimensão da carteira apresentar estabilidade suficiente, mais do que angariar o maior número de clientes possível, as empresas têm como objectivo angariar clientes rentáveis, mesmo que isso implique um menor crescimento em termos de quota de mercado. Reter apenas os clientes com menor risco associado reflecte-se, a longo prazo, numa rentabilidade mais estável.

Com a entrada das seguradoras Directas no mercado o aumento da competitividade foi notório que, agregado à conjuntura actual, se reflectiu numa maior instabilidade nos preços. As empresas vêem-se obrigadas a fazer frequentes descidas e subidas de preços, normalmente associados a campanhas temporárias, para diferenciação de oferta face à concorrência.

As alterações nos preços são visíveis na sucessiva diminuição do prémio médio de seguro ao longo dos anos. Segundo publicação de Gomes, Ana Santos (2010), nos primeiros anos as descidas dos prémios médios estavam relacionados com a descida das taxas de sinistralidade. Completando esta teoria, Rita Sambado, directora de marketing da Fidelidade Mundial enunciada por Gomes, Ana Santos (2012), refere que outro motivo que pode justificar esta descida do prémio médio, aliada à diminuição do número de coberturas facultativas subscritas é o envelhecimento do parque automóvel português, uma vez que as necessidades e possibilidades de subscrição se alteram com a idade do veículo.

Contudo a justificação dada pelo Instituto de Seguros de Portugal (ISP), para a diminuição do prémio médio de seguro automóvel nos últimos anos, tem como fundamento a conjuntura económica actual e o acréscimo dos níveis de competitividade. Informação reforçada por Dias, Mónica (2014) que declarou que, "Até há uns anos, eram as seguradoras que ditavam o preço, mas hoje em dia já não é assim e isso deve-se à maior concorrência, à crise e às próprias pessoas que regateiam cada vez mais os preços".

De acordo com dados da APS o prémio médio de seguro automóvel entre 2003 e 2013 sofreu uma quebra de aproximadamente 30%, passando de 335€ para 224€, apresentado uma descida de cerca de 100€ por apólice.

Esta diminuição do valor dos prémios tem grandes repercussões na estabilidade e rentabilidade das seguradoras, pois os serviços continuam a ser prestados com a mesma ou melhor qualidade, (resultado visível na diminuição do tempo médio de resolução de sinistros), mas a entrada de capital para fazer face às responsabilidades para com os segurados diminuiu, obrigando as seguradoras a reduzir custos administrativos e a esmagar as margens de lucro que até então obtinham.

As Seguradoras Directas representam actualmente uma forte ameaça para as Seguradoras Tradicionais, no entanto a sua quota de mercado continua a representar uma pequena parcela do peso total da quota de mercado do seguro automóvel português, mesmo que o seu crescimento tenha sido significativo ao longo dos anos.

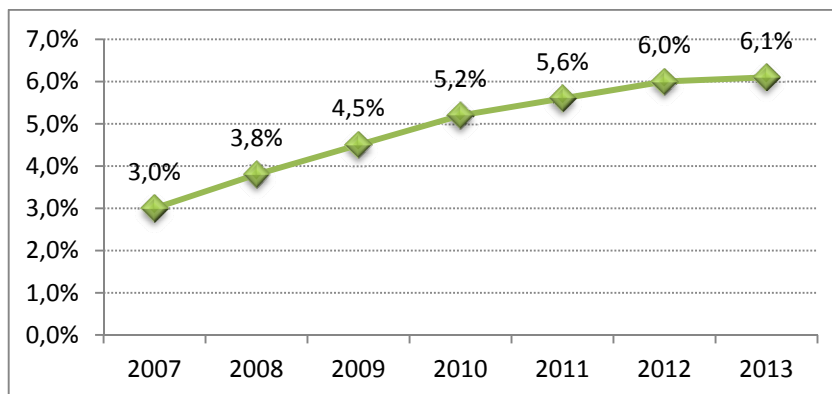


Figura 2.2 - Quota de mercado das Seguradoras Directas em Portugal em PBEs (2007-2013)

Até 2007 apenas duas seguradoras formavam o mercado das Directas em Portugal, a Okteleseguros e a Seguro Directo (recentemente alterada para Direct). Após essa data outras seguradoras iniciaram a sua actividade, como é o caso da N Seguros e da LOGO e mais recentemente a Ipronto e a Seguros Continente.

De acordo com dados da APS, o conjunto das Seguradoras Directas no final de 2007 representava apenas 3% da quota de mercado (em PBEs) no seguro automóvel em Portugal. Como se pode observar na Figura 2.2 o seu crescimento foi bastante rápido nos anos seguintes, abrandado o seu ritmo no último ano, onde o aumento foi de apenas 0,1 ponto percentual, fixando-se no final do ano de 2013 em 6,1%.

O abrandamento deste crescimento, segundo dados apresentados por Melo, Catarina (2014) no site do jornal Diário Económico, pode estar relacionado com o facto do prémio médio das Seguradoras Tradicionais se encontrar cada vez mais próximo dos prémios praticados pelas Seguradoras Directas, também apelidadas como “low-cost”. Esta alteração de prémios de seguro pode ser em parte, justificado através da declaração de Pedro Seixas Vale em entrevista ao jornal Diário Económico, onde afirma que “O enquadramento macroeconómico é pouco estimulante para a actividade seguradora”, obrigando as empresas a praticar preços mais apelativos que a concorrência.

Tendo por base a conjuntura actual e queda do mercado de seguro automóvel, uma das apostas das seguradoras para combater a diminuição de rentabilidade é o alargamento da oferta, diversificando-a ao nível produtos e serviços de modo a captar

mais clientes. No mercado das Directas tem-se verificado um investimento mais acentuado em novos produtos multirriscos habitação e saúde (continuando a actuar no ramo não vida), de forma a captar mais clientes e fidelizar os que compõem a carteira.

Antes de se fazerem investimentos desta natureza é necessário garantir que a empresa disponha de capital suficiente e que a gestão do actual risco esteja a ser feita de modo a que a entrada de novos riscos não comprometa a solvência da seguradora. Contudo, a aposta em novos ramos de actividade também pode ser encarada como uma forma de diversificar o risco, tornando-o menor uma vez que o ramo automóvel apresenta níveis de risco bastante mais elevados que os restantes ramos referidos.

2.3. SINISTROS

2.3.1. Tipos de danos

Existem vários tipos de danos que podem decorrer de um sinistro automóvel. Existem os danos corporais que afectam a vida, saúde ou integridade física de uma pessoa, sendo os danos com maiores custos para a seguradora, dados os custos hospitalares, de internamento ou indemnizações devido a danos irreversíveis na saúde do lesado, ou mesmo, pela perda da vida.

Para além dos danos corporais, existem os danos materiais onde estão contemplados os danos que afectam bens materiais de forma parcial ou total. Este tipo de dano é mais frequente, mas os seus custos individuais normalmente são menores, não causando tantos prejuízos à seguradora a nível unitário. Porém, dada a sua maior frequência, os custos associados à totalidade dos danos continuam a ser elevados. Nos danos materiais o pior cenário, em termos de custos para a seguradora, ocorre perante uma perda total da viatura. Tal verificam-se quando os danos causados na viatura são de tal modo gravosos que não podem, ou não devem, ser reparados por existir a possibilidade de porem em causa questões de segurança. Considera-se perda total quando:

- O veículo desapareceu ou ficou totalmente destruído;
- O veículo sofreu danos impossíveis de reparar.

Quando é declarada perda total, a indemnização a pagar ao cliente corresponde:

- Ao valor venal do veículo (valor antes do sinistro) deduzido do valor do salvado (valor após o sinistro), se o salvado ficar na posse do proprietário;
- Ao valor venal da viatura se o veiculo ficar na posse da seguradora.

2.3.2. Cenário em Portugal

Como já referido, o transporte particular teve um aumento significativo nas últimas décadas, dando origem a um maior fluxo rodoviário nas estradas. Com ele o número de sinistros registou um aumento significativo, que se foi agravando de ano para ano, atingindo aproximadamente 930 milhares de sinistros abertos em 2010. Após essa data o número de sinistros começou a diminuir fixando-se nos 780 milhares no final do ano de 2012.

Os sinistros, na sua grande maioria, provocam danos materiais, sendo a parcela de danos corporais bem mais pequena. Contudo, os danos corporais são por norma mais gravosos, acarretando perdas mais significativas e custos mais elevados.

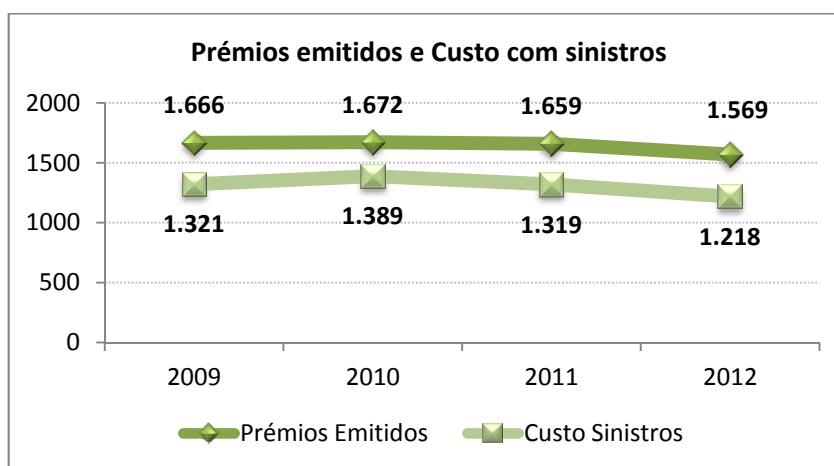


Figura 2.3 - Prémios Emitidos e Custos com sinistros entre 2009 e 2012 (milhares)

Do mesmo modo que o número de sinistros registou aumentos de ano para ano, também os custos com sinistros aumentaram. O valor gasto pelas seguradoras com o pagamento dos sinistros apresentou um crescimento até 2010, diminuindo nos anos seguintes, como se pode observar na Figura 2.3. A redução dos custos totais com sinistros está relacionada principalmente com menores custos com danos corporais ao longo dos anos. Por sua vez, os custos com danos materiais têm sido mais elevados, impedindo que a redução dos custos na sua globalidade seja maior.

Uma parcela significativa de feridos graves e até mesmo de mortes resultam de acidentes rodoviários, que para além do imensurável valor das perdas humanas, estes sinistros representam grandes despesas para as companhias de seguros e para a economia do país. O número de vítimas por acidentes rodoviários apresenta uma evolução decrescente, registando-se diminuições notórias nas vítimas mortais e feridos graves, como se pode observar na Figura 2.4. Segundo estudo efectuado por

Guerreiro (2008), a diminuição da sinistralidade em Portugal foi motivada pelo aumento considerável de auto-estradas, que apresentam melhores condições de circulação, e por medidas de segurança aplicadas ao longo dos anos.

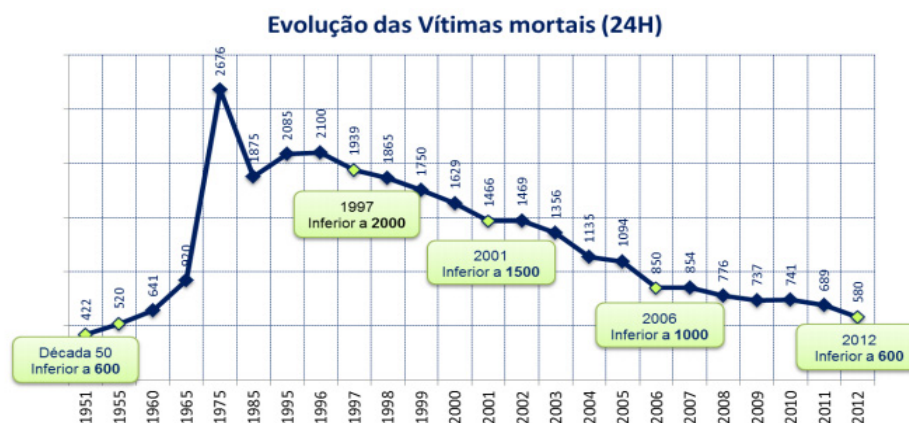


Figura 2.4 - Evolução das Vítimas mortais (1951-2012)

Entre algumas das medidas que foram tomadas para combater o elevado número de vítimas mortais nas estradas destacam-se a inspeção periódica obrigatória, em 1992, a fim de diminuir os riscos causados por mau estado das viaturas; a obrigatoriedade do uso do cinto de segurança; redução do limite de velocidade dentro das localidades; uso de capacete, em 1993, diminuindo o risco de ferimentos graves ou mortes por falta de segurança; agravamento das sanções em 2001 e 2005 e a criação de campanhas de sensibilização e prevenção rodoviária, tanto para condutores como para peões. É de realçar que mesmo com todas as medidas apresentadas muitos dos acidentes têm como causa o não cumprimento dos limites de velocidade e a falta de cuidados de segurança dos condutores e ocupantes.

De acordo com valores apresentados pela Autoridade Nacional de Segurança Rodoviária (ANSR) em 2013 o total de acidentes com vítimas apresentou um aumento de 1,6% face a 2012, contudo o número de vítimas mortais continua a diminuir, registando um valor 9,6% inferior ao ano anterior, o que se reflecte num índice de gravidade de 1,7 em 2013 (0,2 pontos inferior a 2012).

Relacionando os níveis de sinistralidade com regiões verifica-se que os sinistros são mais frequentes em Lisboa e Vale do Tejo e região centro, facto este que se justifica com a densidade populacional da região. Quanto à natureza dos acidentes em Portugal, a colisão abrange cerca de metade dos acidentes com vítimas, sendo a colisão frontal responsável pelo maior número de vítimas mortais.

Mesmo com a evolução apresentada ao nível do número de sinistros e seus custos, os danos corporais continuam a ser uma preocupação para as seguradoras. As despesas com este tipo de danos ainda representam uma fracção elevada no total de custos com sinistros. Custos estes que se mantêm bastante elevados face aos prémios emitidos, podendo pôr em causa os níveis de rentabilidade da empresa.

2.4. Risco

Citando Corte Real (2009), pode definir-se risco como: “Risco é a probabilidade de perda”. Como tal, uma gestão adequada do risco permite às empresas identificar oportunidades existentes no mercado, evitar perdas inesperadas, aumentar a estabilidade e os resultados, ampliando o retorno dos investimentos e criando valor para as mesmas.

Porém, existem riscos que as seguradoras conhecem mas que o controlo não está ao seu alcance. As constantes alterações económicas e financeiras do país, que se reflectem em variações das taxas de juro e impostos, bem como os acidentes naturais ou até mesmo o aumento da concorrência têm reflexos na sua actividade, podendo diminuir a rentabilidade da empresa. Como tal, as seguradoras devem dispor de meios que lhes permitam minimizar os impactos de acontecimentos externos no seu negócio.

Existem três componentes fundamentais que compõem os riscos a que as seguradoras estão expostas, a volatilidade (como o aumento ou diminuição súbita da carteira, alterando o actual nível de risco), a incerteza (devido a modelos mal calculados ou parâmetros mal estimados que possam originar valores não esperados) e a existência de eventos externos (pouco frequentes, difíceis de modelar e que podem originar grandes custos).

Para além dos riscos relacionados com a actividade seguradora – risco de prémios e risco de reservas - as companhias de seguros estão expostas a outros tipos de riscos que influenciam toda a sua actividade. Estes devem ser contemplados nos cálculos dos requisitos de capital que a seguradora deve deter de modo a evitar perdas maiores. Os riscos referidos podem ser enquadrados em Risco de mercado, Risco de contraparte e Risco operacional. A descrição mais detalhada dos tipos de risco pode ser consultada na secção Anexos (9.1.1.2. Tipos de Risco).

Em contrapartida, as companhias de seguros têm menor risco de liquidez que o restante comércio, pois usufruem de um ciclo de produção invertido. Primeiro adquirem os prémios dos clientes e posteriormente, caso o cliente tenha um sinistro, prestam os seus serviços.

O maior risco das seguradoras centra-se no seu passivo, que pode não estar directamente dependente da procura. Sendo o seguro automóvel de responsabilidade civil obrigatório por lei, a procura reúne todo o tipo de consumidores, potencialmente mais ou menos rentáveis, originando um carteira com diferentes níveis de risco, tanto em termos de quantidades como de gravidade.

As perdas causadas por um pequeno grupo de sinistrados devem ser cobertas pelos prémios adquiridos de um grande número de não sinistrados, aplicando-se deste modo a lei dos grandes números, mas para tal é necessário que os riscos existentes sejam independentes e homogéneos. De modo a combater situações que gerem grandes perdas, a companhia de seguros tem de garantir que a tarifa aplicada esteja calculada de acordo com os riscos que assume e que permita o cumprimento de todas as suas responsabilidades, como seguradora para com os clientes e como empresa para os accionistas.

No que respeita ao seguro automóvel, os custos com sinistros continuam a absorver grande parte do valor que as seguradoras cobram ao total dos seus clientes pelos riscos que assumem. Como se pode observar na Figura 2.5, e segundo dados extrapolados de sinistralidade disponibilizados pela APS, os custos com sinistros consomem, há vários anos, a cerca de 80% dos prémios emitidos, estando no entanto este valor a diminuir nos últimos anos, fixando-se no final do ano de 2012 nos 75,5%.

Como referido anteriormente, os custos com sinistros apresentam uma evolução decrescente mas, conjuntamente com estes, os prémios emitidos também têm diminuído, impedindo uma maior acumulação de capital por parte das seguradoras.

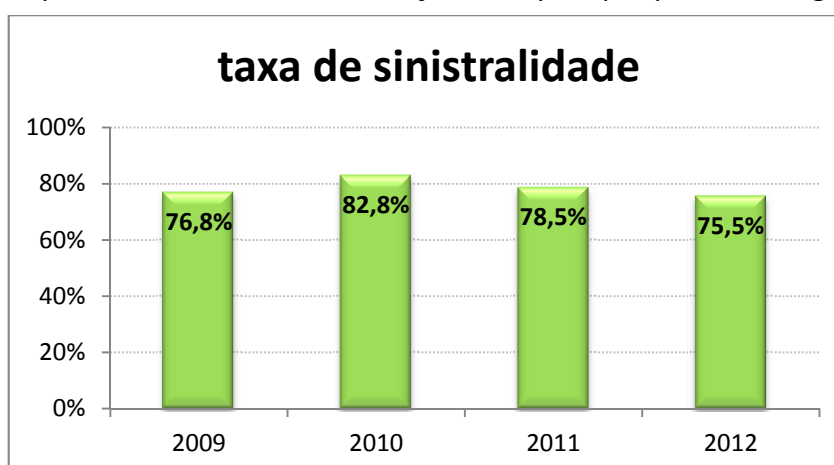


Figura 2.5 - Taxa de sinistralidade entre 2009 e 2012

Esta realidade demonstra a necessidade cada vez maior de conhecer a rentabilidade dos clientes para o cálculo de uma tarifa bem ajustada ao risco, pois os

sinistros não correspondem à totalidade dos riscos a que a seguradora se encontra exposta, nem à totalidade dos custos da mesma. O conjunto de todos os custos face aos ganhos pode influenciar a estabilidade da rentabilidade da empresa ou até mesmo pôr em causa a solvência da mesma.

2.5. SOLVÊNCIA

Uma seguradora, enquanto empresa, encontra-se exposta à volatilidade financeira do país e, de acordo com a actual evolução e instabilidade dos mercados, conhecer a rentabilidade dos clientes é fundamental para a construção de uma tarifa adequada a cada segmento de mercado de acordo com o seu perfil de risco. Uma tarifa bem calculada permite à empresa o cumprimento das responsabilidades enquanto seguradora, dando resposta a todos os riscos sejam eles esperados ou inesperados.

A função de uma empresa de seguros é, mediante o pagamento do prémio de seguro, em caso de sinistro assumir os custos associados aos contratos que subscreve durante o seu período de vigência. De acordo com o apresentado, as responsabilidades de uma seguradora estão muito relacionadas com os índices de sinistralidade, uma vez que grande parte das provisões técnicas é composta pelas provisões para prémios não adquiridos, para riscos em curso e para sinistros.

Para além dos riscos que assume perante os segurados, uma companhia de seguros encontra-se exposta a outros riscos, e todos eles, em conjunto com um conhecimento e gestão adequados da carteira, também necessitam de ser medidos, geridos e controlados. No sistema de solvência que actualmente vigora, o nível de solvência das seguradoras apenas depende do montante de prémios ou sinistros, não existindo uma relação entre os requisitos de capital e o conjunto de riscos a que a empresa se encontra exposta, impedindo uma correcta gestão de risco. Este facto pode originar uma deficiente gestão do negócio, prejudicando os níveis de rentabilidade da empresa devido aos restantes custos não contabilizados.

O projecto de Solvência II tem como objectivo a criação de um sistema de solvência coerente baseado no risco, que permita captar todos os riscos associados ao mercado dos seguros. Este projecto possibilita às empresas a criação dos próprios modelos internos de solvência de acordo com os reais riscos assumidos pela companhia, incentivando a empresa a avaliar e gerir de modo mais autónomo e eficiente os seus próprios riscos.

A existência de modelos internos capacita as empresas a um controlo mais apertado dos riscos, diminuindo a probabilidade de insolvência. Permite também um correcto cálculo dos requisitos de capital tanto para sinistros como para cobrir os possíveis riscos de mercado, incumprimento da contraparte, operacional e de subscrição a que a seguradora se encontra exposta.

Ao fazer uma correcta gestão dos seus riscos, reflectindo-a numa tarifa bem calculada, a companhia de seguros está também a construir meios para fazer frente aos desafios provenientes da globalização e, mais ainda, da concorrência. Não se encontra tão vulnerável às alterações do mercado, tornando-se financeiramente mais estável, o que pode originar maior acumulação de capital e aumento de lucros para a empresa.

2.5.1. Solvência II

O novo sistema de solvência, para além de questões quantitativas, pretende proporcionar às autoridades de supervisão meios que lhes permitam verificar antecipadamente as falhas que possam surgir ou detectar rapidamente as que existam. Assim, será possível agir de modo acertado e também de uma forma clara, transparente e coerente com as necessidades da companhia de seguros, evitando perdas que possam ser controladas e combatidas com a devida antecedência.

Uma das alterações visíveis no Solvencia II foi a inclusão de dois novos conceitos dentro dos requisitos de capital - SCR (Solvency Capital Requirement) e MCR (Minimum Capital Requirement), acrescentados às provisões técnicas já contempladas no balanço das seguradoras, proporcionando às empresas uma gestão mais eficaz dos requisitos necessários.

O projecto assenta em três pilares que suportam a solvência das seguradoras:

- O primeiro pilar tem por base com os requisitos quantitativos de capital, que devem representar o real risco a que a empresa está exposta. Este pilar inclui a avaliação de activos, provisões técnicas e capital.
- O segundo pilar tem como função a revisão da adequabilidade dos requisitos de capital calculados ao risco existente. A supervisão deve actuar assim que os níveis de capital se aproximarem demasiado dos níveis definidos para o MCR (Minimum Capital Requirement), de modo a prever e evitar perdas que levem à insolvência da companhia.
- Por último, o terceiro pilar denomina-se por *Conduta de Mercado* e tem como objectivo definir os resultados que serão divulgados a nível público ou

meramente institucional, fumentando uma maior transparência e disciplina de mercado.

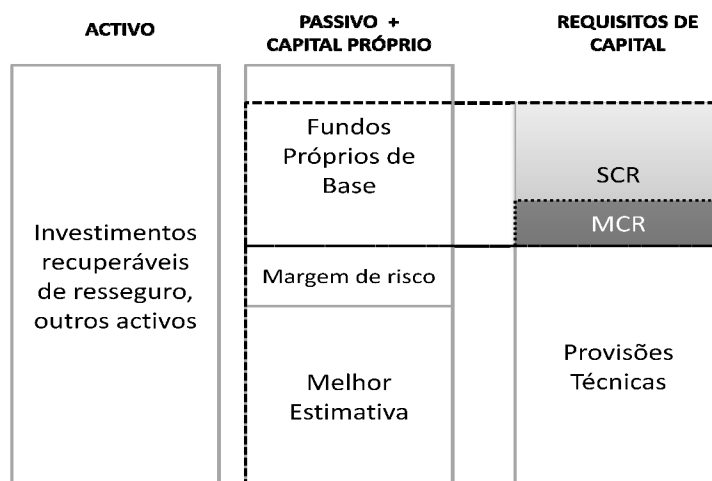


Figura 2.6 - Estrutura do Balanço

O Cálculo do SCR reflecte o nível ideal de requisitos que a empresa deve deter para cumprimento das suas responsabilidades. O seu cálculo pode ser efectuado de acordo com modelos *standard* ou através de modelos internos. A construção de modelos internos utiliza como base o histórico da empresa, sendo reflexo dos níveis de risco assumidos pela companhia num dado momento, e projecta-os para o futuro através de métodos estatísticos.

A conjugação de um conhecimento profundo da rentabilidade da carteira com uma gestão adequada de todos os riscos influencia os níveis dos requisitos necessários, o não tê-los em consideração pode levar a um cálculo de requisitos não adequado, podendo promover uma má gestão de risco e consequentemente um agravamento dos níveis de rentabilidade da empresa.

2.6. RENTABILIDADE

Como referiu Peter L. Bernstein, citado por Corte Real (2009): “A ideia revolucionária que define a fronteira entre os tempos modernos e o passado é o domínio do risco: a noção de que o futuro é mais do que um capricho dos Deuses, e que homens e mulheres não são passivos perante a Natureza”.

Cada vez mais, e de acordo com os valores apresentados sobre a taxa de sinistralidade e prémios emitidos, torna-se uma tarefa árdua obter ganhos muito elevados de rentabilidade. De acordo com esta realidade, há uma grande necessidade, por parte das companhias de seguros, de investir recursos numa boa gestão de risco.

Mas, para tal, não basta ter em atenção os riscos associados às responsabilidades que a empresa assume para com os clientes, há que ter em consideração todos os restantes riscos já referidos, bem como os gastos que a companhia tem com a aquisição e manutenção dos contratos de seguros. Tão ou mais importante que angariar bons clientes é retê-los, pois uma boa carteira pode ser o segredo para uma boa rentabilidade.

A informação sobre a rentabilidade dos clientes pode ser vista como uma oportunidade de maximizar os resultados da empresa e ganhar vantagens competitivas face à concorrência. Com a finalidade de aumentar a rentabilidade dos clientes, Kaplan e Cooper, referidos por Leitão (2002) defendem que se deve:

- Proteger os actuais clientes altamente lucrativos;
- Redefinir os preços de serviços caros, com base no custo do serviço;
- Oferecer descontos, se necessário, para fechar contratos com clientes de baixo custo;
- Negociar relacionamento, proveitosos para todas as partes, que reduzam o custo da angariação dos clientes;
- Deixar que os concorrentes conquistem os clientes que tragam prejuízos constantes;
- Tentar capturar clientes altamente rentáveis dos concorrentes.

Torna-se mais claro perceber a importância da rentabilidade dos clientes quando Foster, enunciado por Leitão (2007), afirma que “ A análise da rentabilidade dos clientes pode resumir-se ao simples facto de que cada dólar de receita gerada pelos clientes não contribui igualmente para novo lucro. Diferentes rentabilidades de clientes podem resultar em diferentes receitas ou diferentes custos”.

Para que a rentabilidade seja maximizada, há que conjugar bons clientes com uma tarifa ajustada ao risco, a fim de que os ganhos gerados sejam maiores que os custos. Uma tarifa bem ajustada ao risco pode levar a uma diminuição da aquisição de novos clientes ou mesmo o abandono dos actuais clientes. Mas por vezes há que sacrificar o crescimento em número de clientes por uma maior diminuição de custos. Ter muitos clientes só é vantajoso se forem clientes rentáveis, caso contrário o crescimento pode originar uma diminuição do nível de rentabilidade devido à aquisição de clientes que tenham associado um elevado risco e que gerem grandes perdas.

2.7. TARIFAÇÃO

A actividade de uma seguradora centra-se na comercialização de contratos de seguro que garantem, mediante pagamento de um prémio de seguro, a reparação ou pagamento de danos causados por sinistros da responsabilidade do segurado. E, de acordo com esta realidade, o grande desafio que se coloca a estas empresas é o cálculo do prémio de seguro a cobrar ao cliente – a tarifa. Os prémios de seguro devem ser suficientes para dar resposta às responsabilidades da empresa, entre as quais cobrir custos futuros de eventuais sinistros. Mas os valores cobrados não podem ser demasiado elevados, estando ajustados ao perfil de risco do segurado de modo a evitar perda de clientes.

Os dados históricos de sinistralidade da empresa são utilizados como base para o cálculo da tarifa, porém há que ter em conta que diferentes clientes se enquadram em diferentes perfis de risco, que devem ser tidos em conta para evitar situações de anti-selecção. Caso a tarifa resultem prémios muitos elevados, estes podem restringir demasiado a entrada de novos clientes, por outro lado, prémios muito baixo promovem a entrada de clientes que apresentem perfis de elevado risco. Um cálculo adequado dos prémios, para além de permitir um equilíbrio de requisitos de capital, pode também ser utilizado com vantagem sobre a concorrência.

A construção de uma tarifa tem em consideração vários factores que caracterizam o risco a assegurar, e que se agrupam de acordo com a sua homogeneidade, criando os escalões tarifários com níveis de maior ou menor risco. A cada escalão é associado um prémio diferenciado em função da gravidade do risco – apólices que se enquadrem no mesmo nível de risco apresentarão prémios semelhantes.

No seguro automóvel existem os factores que caracterizam o objecto seguro (tipo de veículo, Idade do veículo, tipo de utilização, marca, modelo, versão, extras) e o segurado (Idade, género, anos de carta, anos de seguro, código postal). Existem ainda outros factores que são tidos em conta e que influenciam o prémio a cobrar ao cliente, como é o caso do nível de protecção desejado (RC ou DP e coberturas subscritas) e fraccionamento do pagamento do prémio de seguro, que podem originar agravamentos ou descontos face ao prémio resultante dos factores anteriormente referidos.

O peso que cada variável tem no cálculo da tarifa é definido de acordo com dados históricos de sinistralidade. Estes pesos devem ser revistos com regularidade pois os níveis de sinistralidade estão em constante mudança e todos os anos surgem

novos veículos, com diferentes características. Estas alterações devem ser reflectidas no cálculo dos pesos das variáveis de tarifação, pois é aconselhável que o modelo se encontre actualizado de acordo com fundamentos reais e pressupostos actuariais e económicos que descrevam a realidade actual. Prémios devidamente calculados são fundamentais para a optimização dos resultados da seguradora.

O valor estimado pela empresa de seguros, para cobrir todos os custos associados à regularização de sinistros, assume uma grande importância para o cálculo dos requisitos de capital necessários para a seguradora cumprir com as responsabilidades assumidas. A incerteza deve estar contabilizada no cálculo da melhor estimativa, bem como o grau de protecção pretendido e a perspectiva de continuidade do negócio. Porém, não é conveniente que o valor para imprevistos seja muito elevado de modo a não reter capital em excesso que poderia ser utilizado para outros investimentos.

Grande parte destes requisitos está definido num conjunto de medidas que foram formalizadas no Projecto Solvência II que, tal como referido anteriormente, tem como principal objectivo desenvolver um sistema que permita determinar o montante de capital que cada seguradora, no decorrer das suas actividades, deve deter para que consiga cumprir com os seus deveres para com os segurados, bem como para fazer face a custos que não estejam previstos, de forma rentável não comprometendo a solvência da mesma.

3. OBJECTO DE ESTUDO

3.1. DESCRIÇÃO DA EMPRESA

A Seguros Logo, SA é uma seguradora Directa, detida a 100% pela companhia de seguros Tranquilidade. A Seguros Logo iniciou a sua actividade em Janeiro de 2008, apenas no ramo automóvel, delimitando comércio a seguros para automóveis ligeiros de passageiros. Em Novembro de 2009, alargou a oferta a ligeiros comerciais de modo a não restringir tanto a sua carteira, uma vez que este tipo de veículos tem vindo a ganhar peso no mundo automóvel, mesmo para utilização pessoal. Posteriormente, lançou o Logo Casa no final de 2009 e o Logo Saúde em 2012, continuando, no entanto, o Logo Auto a ser o seu principal produto.

Os seus produtos são comercializados com recurso a dois canais de distribuição, telefone e Internet, sendo esta uma das razões que realça a sua característica *low cost*. Desta forma, apresenta menores despesas com comissões, colaboradores e alugueres, traduzindo-se num menor custo por contrato.

Na apresentação dos resultados será mencionado um terceiro canal (Parceiros), onde estão representadas as parcerias que a empresa foi realizando ao longo da sua actividade, mas que têm por base o canal de distribuição Internet.

A empresa reforça a sua imagem inovadora apresentando uma visão de simplicidade e rapidez de serviços, pois com o auxílio das novas tecnologias é possível recolher os dados de cada cliente sem necessidade de grandes burocracias, tornando os processos mais simples e rápidos, permitindo a subscrição imediata do contrato de seguro.

3.2. RAZÕES PELA ESCOLHA DE UM CANAL DIRECTO

Dada a actual situação económica da sociedade portuguesa e a variedade de oferta por parte das diversas empresas, existem cada vez mais clientes "*price-seekers*" que, para obterem o melhor produto a baixo preço, gastam algum do seu tempo a estudar a oferta de mercado o que, neste contexto, se enquadra com os clientes que apenas pretendem cumprir a lei, procurando o seguro mais barato, não valorizando tanto com a protecção do seu veículo.

Para que as empresas consigam atrair este tipo de clientes têm de adaptar a sua oferta de forma a conjugarem preços baixos e serviços rápidos e de qualidade. Esta procura é muitas vezes efectuada através da Internet, onde há maior difusão de

informação e a pesquisa é feita de forma simples, autónoma e rápida, sem qualquer tipo de pressão por parte de mediadores.

As novas tecnologias fazem, cada vez mais, parte do quotidiano dos portugueses. Muitos são os que alteraram os seus hábitos de consumo dando preferência à procura e aquisição de bens e serviços através da Internet e telefone, por valorizarem a comodidade, simplicidade e rapidez de serviços, não sendo sujeitos a tantas limitações de tempo e local.

Enquadrando esta realidade com o universo dos seguros, para além de comparação de preços, também é possível subscrever e activar um seguro em qualquer local com acesso à Internet ou telefone. Para o cliente é mais rápido e cómodo e para a empresa torna-se mais fácil promover a marca e o produto a um maior número de consumidores, com custos mais baixos de exploração.

3.3. PERFIL DO CLIENTE ALVO:

A Logo procura atingir os consumidores residentes nas grandes zonas urbanas como Lisboa e Porto, onde existe uma maior concentração populacional e onde se verifica uma maior concentração de automóveis. A Logo aposta, preferencialmente, em clientes com idades compreendidas entre os 25 e 45 anos, uma vez que consumidores com idade inferior a 25 anos têm associado um risco bastante elevado devido à pouca experiência e consumidores acima dos 45 anos não serão, actualmente, utilizadores assíduos do comércio electrónico ou por telefone.

Grande parte dos clientes alvo já possui seguro automóvel, pois este é um seguro obrigatório por lei e, como tal, uma das características necessárias para que a Seguros Logo consiga atingir estes clientes é a sua aceitação à mudança, ou seja, clientes que, perante uma proposta mais atractiva (melhores serviços ou melhor preço), aceitem mudar de seguradora.

3.4. ESTRATÉGIA DE NEGÓCIO

Num período inicial, uma seguradora tem como principal foco a angariação de clientes, de modo a construir uma carteira com uma dimensão suficiente para se afirmar no mercado das seguradoras Directas e obter determinada quota de mercado.

Anos depois do início da sua actividade, a empresa, já possuindo um número de clientes que considere razoável, e não deixando de ter como objectivo expandir a sua

operação, pode considerar uma reorientação da estratégia do negócio, procurando aumentar a sua rentabilidade e, conseqüentemente melhorar os níveis de rentabilidade. Uma das formas de melhoria passa pela redução de custos com sinistros, que pode resultar de ajustes da tarifa, a fim de constituir uma carteira de clientes com perfis associados a baixos níveis de sinistralidade.

Uma das questões com bastante relevância para a gestão do negócio é a definição dos clientes mais e menos rentáveis. Uma grande carteira só é um bom activo se gerar rentabilidade consistentemente. Após uma análise cuidada da rentabilidade dos seus clientes, tendo por base alguns parâmetros de avaliação considerados relevantes e que descrevam claramente os clientes, a empresa poderá aplicar medidas mais adequadas para angariar e reter clientes com um determinado perfil e criar restrições à entrada dos que considera terem fraca rentabilidade. Perspectivar com maior fiabilidade a rentabilidade de cada cliente é crucial para a constituição de uma carteira que associa dimensão comercial e estatística ao potencial de gerar a rentabilidade esperada pelos accionistas e manter-se solvente, assegurando assim os interesses dos seus clientes.

Só assim a Seguro Logo conseguirá um volume de negócio com dimensão que, associado à solidez técnica resultante de uma cuidada gestão de risco, lhe permita impor-se no mercado de seguro automóvel a nível global com alguma segurança.

Este estudo pretende colaborar para este objectivo, apresentando uma análise estatística da carteira, bem como a identificação de características de um “perfil” de segurado rentável.

3.5. OFERTA LOGO

Todas as análises efectuadas no decorrer deste projecto incidirão apenas sobre o produto Logo Auto. A oferta automóvel é composta por várias opções - Logo Moto, Logo Light, Logo Topping e Logo Safe dentro dos produtos de Responsabilidade Civil e Logo Max de Danos Próprios.

Na Tabela 3.1 é possível conhecer a oferta da seguradora em termos do produto Auto e as coberturas que compõem cada produto. Descrição mais detalhada dos produtos na secção Anexos (9.1.1.3 - Oferta Logo).

Produtos Coberturas	<u>LOGOMoto</u>	<u>LOGOLight</u>	<u>LOGOTopping</u>	<u>LOGOSafe</u>	<u>LOGOMax</u>
	Resp. Civil "contra terceiros"	Resp. Civil "contra terceiros"	Resp. Civil "contra terceiros"	Resp. Civil "contra terceiros"	Danos Próprios "contra todos"
Responsabilidade Civil	Sim	Sim	Sim	Sim	Sim
Assistência em Viagem	Sim	Light/Normal	Normal	Normal/VIP	Normal/VIP
Protecção Jurídica	Sim	Sim	Sim	Sim	Sim
Protecção dos Ocupantes	-	-	Base/Extra	Base/Extra	Base/Extra
Quebra Isolada de Vidros	-	-	(opcional)	Sim	Sim
Choque, Colisão e Capotamento	-	-	-	-	Sim
Incêndio, Raio ou Explosão	-	-	-	-	Sim
Furto ou Roubo	-	-	-	Sim	Sim
Actos de Vandalismo	-	-	-	-	(opcional)
Fenómenos da Natureza	-	-	-	-	(opcional)
Veículo de Substituição	-	-	(opcional)	(opcional)	(opcional)
Indemnização Extra	-	-	(opcional)	(opcional)	(opcional)

Tabela 3.1 - Oferta LOGO; Produtos e Coberturas

4. MODELOS LINEARES GENERALIZADOS

Uma correcta gestão da carteira de clientes é uma preocupação constante nas empresas e, no caso das seguradoras, a análise da rentabilidade dos clientes e da sua relação com a sinistralidade são pontos fundamentais para uma gestão saudável.

Nos anos 60, para analisar a sinistralidade e os resultados da sua actividade, os actuários desenvolveram uma técnica formada por um conjunto de equações que tinham por base os dados dos clientes, as variáveis utilizadas para a tarificação e outros parâmetros de risco que a empresa estava a cobrir. Após a criação das equações, as mesmas eram corridas e alteradas iterativamente até se chegar o mais próximo possível da solução óptima. Esta técnica, porém, não possibilita testar a significância de cada variável no resultado, nem o nível de ajustamento do modelo.

Para combater as falhas detectadas na técnica apresentada, contínuos estudos foram efectuados até que, nos anos 70, surgiram os Modelos Lineares Generalizados (MLG) introduzidos por Nelder e Wedderburn. Estes modelos, como o próprio nome indica, têm uma estrutura generalizada, mas também flexível, permitindo o cálculo do erro presente nas estimativas. Os MLG são aplicáveis a problemas onde existe a necessidade de modelar a relação entre variáveis e estudar a influência das mesmas sobre uma outra variável. Quando estes modelos foram desenvolvidos, todo o processo de modelação de dados era demorado, apenas com o desenvolvimento tecnológico a nível informático e estatístico se tornou possível a utilização deste tipo de modelos com relativa rapidez.

Os MLG são uma extensão dos modelos lineares clássicos (MLC), mas de forma mais generalizada, pois a sua metodologia tem por base uma abordagem unificada de muitos procedimentos estatísticos utilizados. Estes modelos podem ser diferenciados de acordo com o tipo de respostas pretendido, as variáveis resposta podem ser de natureza contínua, dicotómica, na forma de proporções ou na forma de contagens. Este último tipo de modelo é frequentemente utilizado quando se pretende analisar a frequência de determinados acontecimentos, como é o caso da análise da sinistralidade.

Uma das vantagens dos MLG é a flexibilidade que apresentam, possibilitando especificar uma função de ligação e a distribuição estatística que melhor descreve as observações, dentro do espectro da Família Exponencial de Distribuições, para além da possibilidade de medição da qualidade do ajustamento do modelo através, por exemplo, da análise dos resíduos.

No caso da tarificação de seguros automóvel, o recurso aos MLG permitiu às seguradoras uma nova forma de modelação da sinistralidade. A utilização deste tipo de modelos permite conhecer os efeitos de cada factor e de cada nível de tarificação de forma mais ajustada e com maior precisão.

O recurso aos MLG para modelação da sinistralidade foi inicialmente utilizado considerando apenas a frequência dos sinistros. Só posteriormente foi inserida na análise da severidade dos mesmos através da análise dos custos resultantes destes permitindo uma análise mais clara do real impacto dos sinistros na gestão da carteira, (Santos, Susete [2008]).

4.1. MODELOS LINEARES CLÁSSICOS

Os modelos lineares clássicos (MLC) são utilizados quando se pretende conhecer o comportamento de uma variável aleatória Y (resposta), em função de um vector $\mathbf{x} = (x_1, \dots, x_k)^T$ de k variáveis explicativas (covariáveis).

A variável Y pode ser de natureza contínua, discreta ou dicotómica. AS covariáveis determinísticas ou estocásticas, podem ser contínuas, discretas, quantitativas de natureza ordinal ou dicotómicas.

Pode assumir-se que os dados apresentam a forma

$$(y_i, x_i), \quad i = 1, \dots, n,$$

resultantes da realização de (Y, \mathbf{x}) em n unidades experimentais, sendo as componentes Y_i do vector aleatório $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ independentes.

Este tipo de modelos pode ser escrito através da expressão:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4.1.1)$$

onde \mathbf{X} corresponde, em geral, à matriz de covariáveis ou variáveis explicativas, com um primeiro vector unitário, associada a um vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ de parâmetros, que se espera que expliquem parte da variabilidade associada a \mathbf{Y} , sendo $\boldsymbol{\varepsilon}$ um vector de erros aleatórios com uma distribuição que se supõe Normal, tal que $E(\varepsilon_i)=0$ e $V(\varepsilon_i) = \sigma^2$.

Pressupõe-se ainda que:

- As observações individuais são independentes e seguem uma distribuição Normal, de média variável e variância constante, $Y_i \sim N(\mu_i, \sigma^2)$;
- A média resulta de uma combinação linear das variáveis explicativas,

$$\mu_i = E(Y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, i = 1, \dots, n$$

- Uma vez que o vector β é estimado e pode apresentar componentes negativas, Y e μ pertençam ao conjunto dos números reais ($Y \in \mathbb{R}$ e $\mu \in \mathbb{R}$)

No entanto, os MLC apresentam algumas limitações, não permitindo a sua aplicação em determinadas situações práticas como, por exemplo, nas situações em que:

- A variável resposta não segue uma distribuição Normal, podendo mesmo assumir uma distribuição que não seja contínua;
- A variável resposta assume apenas valores inteiros ou positivos;
- A média é obrigatoriamente positiva e, pode não ser escrita como combinação linear das variáveis explicativas e dos parâmetros uma vez que através da estimação de β podemos obter valores negativos;
- A variância não é constante, dependendo da média (fenómenos que revelem heterocedasticidade).

Perante as limitações dos MLC, vários estudos foram desenvolvidos para que as mesmas fossem ultrapassadas. Até que, como já referido, nos anos 70 foram propostos os MLG, ultrapassando algumas das barreiras encontradas, como é o caso da distribuição da variável resposta, que pode assumir qualquer distribuição da Família Exponencial, deixando de ter a obrigatoriedade de seguir uma distribuição Normal. A sua média deixa de ser escrita como uma combinação linear e passar a ser função de um preditor linear, sendo este uma combinação linear das variáveis explicativas ($E[Y] = g^{-1}(\eta) = g^{-1}(\sum_i^p X_j \beta_j)$).

A relação entre a média de Y e o preditor linear passar a ser expressa através da função de ligação g , função monótona e diferenciável, tal que $\eta = g(\mu)$. Esta função deve ser escolhida de acordo com o problema em estudo.

4.2. A FAMÍLIA EXPONENCIAL

Resultados acerca da Família Exponencial podem facilmente ser encontrados na literatura. No texto que se segue, utilizou-se a referência Turkmann e Silva (2000).

Nos MLG, a variável resposta pode seguir uma de várias distribuições possíveis dentro da Família Exponencial, o que lhes atribui propriedades específicas. Uma variável aleatória Y tem distribuição pertencente à Família Exponencial se a sua função

densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.p.) se puder escrever na forma:

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}, \quad (4.2.1)$$

sendo θ e ϕ parâmetros escalares, $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ funções reais conhecidas. θ corresponde ao parâmetro canónico que está relacionado com a média e ϕ ao parâmetro de escala que está relacionado com a variância.

Existem ainda condições associadas a esta função:

- $a(\phi)$ é positiva e contínua;
- $b(\theta)$ é diferenciável até à segunda ordem, sendo a segunda derivada uma função positiva;
- $c(y, \phi)$ é independente do parâmetro θ

Como já referido anteriormente, os MLG são uma extensão do modelo linear clássico,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon, i = 1, \dots, n \quad (4.2.2)$$

caracterizando-se essencialmente com base três componentes:

- A variável resposta segue uma distribuição de probabilidade que pertença à Família Exponencial;
- Um preditor linear $\eta = \sum_{i=1}^p X_i \beta_i$;
- Uma função de ligação g tal que $\eta = g(\mu)$.

Relativamente à função de ligação, esta deve ser escolhida de acordo com o problema em estudo.

Cada uma das distribuições da Família Exponencial tem associada a sua função de ligação canónica. No quadro seguinte é possível observar as funções de ligação associadas às distribuições mais comuns.

Normal	Poisson	Binomial	Gama	Gaussiana Inversa
μ	$\ln(\mu)$	$\ln\left(\frac{\mu}{1-\mu}\right)$	$1/\mu$	$1/\mu^2$

Tabela 4.1- Funções de Ligação Canónica mais comuns

Para além dos modelos de regressão linear clássica também fazem parte dos MLG os de regressão logística, de regressão de Poisson, de análise de variância e covariância, log-lineares para tabelas de contingência multidimensional, entre outros.

Ao tentar modelar dados utilizando os MLG, existem etapas fundamentais que devem ser efectuadas.

4.3. FORMULAÇÃO DOS MODELOS

Após a correcta definição do problema em estudo, o ponto de partida para a formulação dos modelos é uma análise cuidada dos dados, a fim de se escolher qual a distribuição que melhor descreve a variável que se pretende estudar (variável resposta ou variável dependente) e as variáveis explicativas (covariáveis ou variáveis independentes). É indispensável ter sempre presente o problema que se está a analisar, de modo a codificar correctamente as variáveis.

A análise da distribuição da variável resposta deve ter em conta a natureza dos dados (continua ou discreta). A distribuição Logística e Gama são adequadas a dados de natureza contínua, por seu lado para dados de natureza discreta são mais apropriadas as distribuições de Poisson ou Binomial. A análise da distribuição da variável resposta também tem como função a identificação de características invulgares ou problemas existentes nos dados como é o caso dos outliers, que devem ser eliminados.

As variáveis explicativas devem ser seleccionadas de acordo com o problema em estudo. Sobre as mesmas devem ser efectuadas análises estatísticas simples, de modo a verificar quais os factores mais relevantes e verificar a existência de dados incoerentes ou tendências que possam enviesar os resultados.

Ainda no contexto da formulação dos modelos não se pode deixar de parte a escolha da função de ligação, que deve ser compatível com a distribuição do erro associado aos dados.

4.3.1. Regressão Logística

Quando se pretende modelar um conjunto de dados a partir de variáveis explicativas contínuas ou binárias e cuja variável resposta é binária, é frequente recorrer-se à regressão logística. O que a distingue claramente a regressão logística é o facto da variável resposta ser categórica, sendo binomialmente distribuída, $Y_i \sim B(1, \pi_i)$. De acordo com Hair et al (2009), referido por Vale (2010), esta regressão

é uma técnica estatística utilizada quando se está perante dois grupos e se pretende obter a probabilidade de que uma observação pertença a um determinado conjunto, em função do comportamento das variáveis dependentes.

Consiste em relacionar, através de um modelo, a variável resposta com factores que influenciem ou não a probabilidade de ocorrência de determinado acontecimento, sendo a sua utilização frequente quando se pretende analisar dados com resposta binária ou dicotómica. Desta forma, na regressão logística, como a variável resposta possui um carácter não-métrico assume-se como sendo uma variável *dummy* (dicotómica ou binária), atribuindo-se o valor 0 na ausência do atributo e 1 na presença do mesmo, estimando a probabilidade de determinado evento ocorrer ou não. No âmbito deste projecto, a regressão logística será utilizada para avaliar a probabilidade de um cliente gerar ou não lucro para a empresa.

De acordo com Turkman e Silva (2000), pressupondo a existência de n variáveis resposta independentes $Y_i \sim B(1, \pi_i)$ ou $Y_i \sim Ber(\pi_i)$, i.e.,

$$f(y_i|\pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}, \quad y_i = 0, 1 \quad (4.3.1)$$

e que a cada individuo i está associado um vector de especificação \mathbf{Z}_i , que resulta do vector de covariáveis \mathbf{X}_i , $i = 1, \dots, n$.

Pertencendo a distribuição Binomial à Família Exponencial temos que,

$$E[Y_i] = \frac{e^{\theta_i}}{1+e^{\theta_i}}.$$

Como $E[Y_i] = \pi_i$ para o modelo binomial, então

$$\theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right).$$

Colocando $\theta_i = \eta_i = \mathbf{z}_i^T \boldsymbol{\beta}$ conclui-se que a função de ligação canónica que faz a associação entre o valor esperado da variável resposta e as covariáveis é a função *logit*. Deste modo a probabilidade de sucesso, $\pi_i = P[Y_i = 1|X = x_i]$, está relacionada com o vector \mathbf{z}_i através de

$$\pi_i = \frac{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{z}_i^T \boldsymbol{\beta})}. \quad (4.3.2)$$

Deste modo, $\text{Logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln(e^{\theta_i}) = \theta_i$ e

$$\text{Logit}(\pi_i) = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i, \quad (4.3.3)$$

situando-se os valores possíveis de π_i no intervalo $[0,1]$.

A principal vantagem da regressão logística é a capacidade de estimação de probabilidades individuais e, para além desse facto, quando se trabalha com esta regressão não se torna necessária a verificação de alguns pressupostos como é o caso da normalidade das variáveis independentes ou a igualdade das matrizes de variância/covariância para os grupos (ausência de heterocedasticidade). Segundo Hair et al. (2009), enunciado por Vale (2010), outras análises que se podem efectuar são a linearidade das relações entre as variáveis e a ausência de multicolinearidade entre as mesmas.

4.3.2. Regressão Gama

A regressão Gama pode ser utilizada em modelos cuja variável resposta é de natureza contínua e positiva, enviesada à direita e permite a existência de valores elevados na cauda direita da distribuição. Segundo Turkman e Silva (2000) é também utilizada quando a variância é crescente com a média ou quando o coeficiente de variação dos dados for aproximadamente constante.

A distribuição Gama pertence à Família Exponencial, da qual fazem parte, entre outras, a distribuição Exponencial ou Qui-quadrado. A utilização da distribuição Gama pode ser uma boa opção quando os dados a descrever não são simetricamente distribuídos, uma vez que é uma distribuição assimétrica, podendo deste modo descrever melhor a distribuição dos dados. Contudo, há sempre a necessidade de testar a sua adequabilidade.

De acordo com as características que descrevem a distribuição Gama, esta pode ser uma distribuição a considerar na modelação de dados de sinistralidade automóvel, tanto ao nível dos custos com sinistros como da rentabilidade de uma apólice.

Uma variável aleatória Y apresenta uma distribuição gama, $Y \sim G(\mu, \phi)$, se a sua f.d.p for expressa por:

$$f_y(y; \mu, \phi) = \frac{\left(\frac{\phi}{\mu}\right)^\phi}{\Gamma(\phi)} y^{\phi-1} \exp\left(-\frac{\phi y}{\mu}\right), y > 0 \quad (4.3.4)$$

com os parâmetros μ e ϕ positivos, em que μ é corresponde à média, ϕ determina a forma da distribuição e a variância de Y é dada por μ^2 / ϕ .

4.4. AJUSTAMENTO DOS MODELOS

Na fase do ajustamento dos modelos é efectuada a estimação dos parâmetros que compõem o modelo, mais propriamente a estimação dos coeficientes β 's associados às covariáveis e do parâmetro de dispersão, no caso de ϕ existir.

A inferência realizada sobre os dados baseia-se essencialmente na verosimilhança, recorrendo-se ao método da máxima verosimilhança para estimar os parâmetros da regressão e efectuar testes de hipóteses, bem como para averiguar a qualidade do ajustamento.

4.4.1. Estimação de β

O parâmetro β é estimado com recurso ao Método de Máxima Verosimilhança, que permite a obtenção de estimadores consistentes, eficientes (com menor variância), não enviesados e com uma distribuição aproximadamente normal. É de realçar que, mesmo sendo uma técnica robusta de estimação, o estimador obtido pode ser enviesado quando a amostra for de pequena dimensão.

Este método é utilizado devido às suas propriedades e também por permitir a aplicação de testes de hipóteses sobre os parâmetros estimados, bem como aferir a qualidade do ajustamento do modelo.

4.4.2. Testes de hipóteses

Quando se considera um grande número de covariáveis, torna-se importante saber qual o modelo que, com menor número de variáveis possível, permite uma boa interpretação do problema em estudo e que possua um bom ajustamento de dados. Depois de seleccionar as variáveis que à partida se imagina que sejam as que melhor explicam a variável resposta há que testar a sua real contribuição, ou seja, se são estatisticamente significativas ou não. No entanto, é possível que existam casos em que as covariáveis não sejam estatisticamente significativas mas, de acordo com o conhecimento do negócio, se opte por mantê-las no modelo ou vice-versa.

Para se testar se um submodelo do modelo original é estatisticamente melhor que o modelo original (contendo todas as variáveis seleccionadas), as estatísticas mais comuns são a *Estatística de Wald*, *Estatística de Wilks* ou *Estatística de Razão de Verosimilhança*, de acordo com Trukman e Silva, (2000).

Assumido o teste de hipóteses sob a forma:

$$H_0: \mathbf{C}\beta = \xi \quad \text{vs} \quad H_1: \mathbf{C}\beta \neq \xi, \quad (4.4.1)$$

onde \mathbf{C} corresponde a uma matriz $q \times p$, com $q \leq p$, de característica completa q e ξ a um vector de dimensão q previamente especificado.

Quando o teste de hipóteses é utilizado para testar a nulidade do vector parâmetro, este assume a forma:

$$H_0: \mathbf{C}\beta_j = 0 \quad \text{vs} \quad H_1: \mathbf{C}\beta_j \neq 0, \quad (4.4.2)$$

para algum j , sendo $q=1$ e $\mathbf{C}=(0,\dots, 0,1,0,\dots,0)$, um vector com todas as componentes nulas excepto a j -ésima que será igual a 1.

Caso exista uma variável policotómica, que possa assumir $r + 1$ valores distintos, é preferível substituí-la por r variáveis dicotómicas que a representem, havendo r parâmetros β 's que lhe estão associados. De modo a aferir se a variável deve ou não fazer parte do modelo é importante testar se os r parâmetros são estatisticamente significativos.

4.4.2.1. Teste de Wald

A *Estatística de Wald*, Turkman e Silva (2000), baseia-se na normalidade assintótica do estimador de máxima verosimilhança $\hat{\beta}$.

Admitindo que a hipótese nula estabelece que $H_0: \mathbf{C}\beta = \xi$, onde \mathbf{C} corresponde a uma matriz $q \times p$, com $q \leq p$, de característica completa q . Sendo $\hat{\beta}$ o estimador de máxima verosimilhança de β , que segue uma distribuição assintótica $N_p(\beta, \mathfrak{S}^{-1}(\hat{\beta}))$, onde $\mathfrak{S}^{-1}(\beta)$ corresponde à matriz de covariâncias. Sendo o vector $\mathbf{C}\hat{\beta}$ uma transformação linear de $\hat{\beta}$ então, de acordo com as propriedades da distribuição normal multivariada,

$$\mathbf{C}\hat{\beta} \overset{a}{\sim} N_p(\mathbf{C}\beta, \mathbf{C}\mathfrak{S}^{-1}(\hat{\beta})\mathbf{C}^T)$$

Consequentemente, sob a hipótese nula, a *Estatística de Wald* é expressa por:

$$W = (\mathbf{C}\hat{\beta} - \xi)^T [\mathbf{C}\mathfrak{S}^{-1}(\hat{\beta})\mathbf{C}^T]^{-1} (\mathbf{C}\hat{\beta} - \xi), \quad (4.4.3)$$

apresentado uma distribuição assintótica de um χ^2 com q graus de liberdade.

Recorrendo ao teste de hipóteses (4.4.2) e designando por σ_{ii} o j -ésimo elemento da diagonal de $\mathfrak{S}^{-1}(\hat{\beta})$, é possível simplificar a *Estatística de Wald* a:

$$W = (\hat{\beta}_j - \beta_j)^T [\sigma_{ii}] (\hat{\beta}_j - \beta_j) \quad \text{Sob } H_0, \quad W = \frac{\hat{\beta}_j^2}{\sigma_{ii}} \overset{a}{\sim} \chi_1^2$$

Deste modo, rejeita-se a hipótese nula, a um nível de confiança de α , se o valor resultante da *Estatística de Wald* for superior ao quantil de probabilidade $1-\alpha$ de um χ_{α}^2 .

Por norma, a *Estatística de Wald* é utilizada para testar hipóteses nulas sobre componentes individuais. Este teste é um recurso frequentemente utilizado, na

comparação de dois modelos, quando se inicia o processo com o modelo que contém todas as covariáveis e se prossegue a análise considerando modelos alternativos pela exclusão de covariáveis.

4.4.2.2. Teste de Razão de Verosimilhança

A *Estatística de Razão de Verosimilhança ou Estatística de Wilks*, ver Turkman e Silva (2000), baseia-se na distribuição assintótica da razão de máxima verosimilhança, sob as hipóteses H_0 e $H_0 \cup H_1$, sendo definida por:

$$\Lambda = -2 \ln \frac{\max_{H_0} L(\boldsymbol{\beta})}{\max_{H_0 \cup H_1} L(\boldsymbol{\beta})} = -2 \{l(\tilde{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}})\} \quad (4.4.4)$$

sendo $\tilde{\boldsymbol{\beta}}$ o estimador de máxima verosimilhança restrito, que corresponde ao valor de $\boldsymbol{\beta}$ que maximiza a verosimilhança, sujeito a restrições impostas pela hipótese $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}$.

O Teorema de Wilks refere que sob determinadas condições de regularidade, a estatística Λ tem, sob H_0 , uma distribuição assintótica χ^2 , onde o número de graus de liberdade é igual à diferença entre o número de parâmetros a estimar sob $H_0 \cup H_1$ (p) e o número de parâmetros a estimar sob H_0 ($p-q$).

Desta forma, sob H_0 ,

$$\Lambda = -2 \{l(\tilde{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}})\} \stackrel{a}{\sim} \chi_q^2$$

De acordo com o Teste de Razão de Verosimilhança rejeita-se a hipótese nula, $H_0: \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}$, a um nível de significância de α , se o valor resultante da estatística Λ for superior ao quantil de probabilidade $1-\alpha$ de um χ_q^2 .

Este tipo de estatística é frequentemente utilizado quando os modelos que se pretende comparar são encaixados (quando um dos modelos é um submodelo do modelo original).

4.5. SELECÇÃO E AVALIAÇÃO DOS MODELOS

Esta fase tem como objectivo encontrar submodelos que apresentem um número moderado de parâmetros que sejam adequados aos dados, permitindo uma boa interpretação do problema. Outro objectivo desta fase é detectar possíveis discrepâncias entre os dados e os valores preditos que podem estar relacionados com a existência de outliers.

4.5.1. Selecção dos modelos

A fim de se identificar as variáveis independentes mais importantes para a construção do melhor modelo, sem que se perca informação relevante para o problema em estudo, recorreu-se aos métodos estatísticos *Stepwise* e *AIC*.

4.5.1.1. Método Stepwise

Para a selecção das variáveis com maior relevância o método mais utilizado é o *Stepwise*, que verifica a existência de multicolinearidade e corrige automaticamente as possíveis distorções causadas pela mesma. Com o objectivo de incluir ou excluir variáveis independentes ao modelo inicial com base no poder discriminatório, sendo a inclusão ou exclusão feita de forma unitária (uma variável de cada vez), o método *stepwise* permite ter a percepção de quais as variáveis altamente correlacionadas com a variável que está a ser incluída ou excluir no modelo. É importante realçar que todas as suposições indicadas apenas são relevantes para análises discriminantes, não sendo para a regressão logística, uma vez que esta é bastante robusta.

Segundo Araújo (2006) o método *Stepwise* é, por norma, utilizado quando se está perante um modelo inicial com um número elevado de variáveis independentes, sendo retiradas as variáveis com menor relevância. Podendo, no entanto, ser utilizado quando se inicia a selecção das variáveis com um modelo nulo e se acrescentam apenas as variáveis com maior significância estatística.

A selecção das variáveis que melhor explicam a variável dependente é feita de forma sequencial. Em cada passo a variável menos significativa é retirada do modelo em análise, ou no caso de se iniciar com o modelo nulo, é adicionada a variável mais significativa, sendo retido apenas um número reduzido de variáveis independentes. O ideal é que o modelo menor seja tão ou mais explicativo que o completo.

De acordo com Turkman e Silva (2000), o método *Stepwise* utiliza como base de comparação o valor dos p-values obtidos através dos testes de razão de verosimilhança de Wilks entre os modelos com inclusão ou exclusão de covariáveis, a fim de se verificar quais as que devem permanecer no modelo final. Este método começa por calcular o p-value obtido através de teste de Wald, e com base nesse valor, escolhe a variável que em primeira análise deve ser excluída (ou incluída) no modelo final. Quanto menor (ou maior) for o valor do p-value obtido, mais (ou menos) importante é a covariável. Após a escolha da covariável, faz-se uma segunda análise ao seu grau de importância através do valor do p-value resultante do teste de razão de verosimilhança entre os dois modelos - o modelo inicial e o modelo com a exclusão (ou

inclusão) da covariável - e de acordo com o resultado opta-se por manter o modelo contendo a variável ou excluir (ou incluir) a mesma.

O método Stepwise pode ser executado recorrendo a duas vertentes distintas, forward stepwise ou backward stepwise. O que as diferencia é o facto de no primeiro método se iniciar a análise com um modelo nulo, sendo depois incluídas as variáveis significativas. No segundo método utiliza-se o processo inverso, inicia-se a análise com o modelo completo e de acordo com a significância das covariáveis opta-se, ou não, pela exclusão de variáveis.

Neste projecto optou-se pela utilização do modelo Backward Stepwise.

4.5.1.2. Backward Stepwise

Como referido no tópico anterior, no processo de selecção do modelo utilizando o método *Backward Stepwise*, as variáveis são retiradas sequencialmente do modelo original (composto por todas as variáveis escolhidas inicialmente).

Inicia-se o processo pelo ajustamento do modelo contendo todas as covariáveis e, de acordo com os resultados do teste de Wald, retira-se a que apresentar maior p-value, sendo essa a menos significativa.

O passo seguinte é, através do teste de razão de verosimilhança, comparar o ajuste do modelo completo com o obtido após a exclusão da variável. Se o p-value resultante do teste de Wilks, for inferior a 0,05 considera-se o modelo com a variável como sendo o melhor modelo, pois sem essa variável o modelo estimado é menos significativo, caso contrário a variável é retirada. Nas etapas seguintes, têm-se o modelo resultante da etapa anterior como sendo o modelo inicial e aplicam-se os mesmos passos. No caso de, na etapa anterior, não se ter removido a variável e ainda existam variáveis para remover, a variável seguinte a ser analisada será a que apresentar maior p-value a seguir à que foi mantida no modelo.

Quando se obtém um modelo em que todas as variáveis sejam significativas segundo o teste de Wald, o processo termina e o modelo final é constituído por todas as variáveis dessa etapa.

4.5.1.3. Método AIC

O critério de informação de Akaike foi desenvolvido em 1972 por Hirotugu Akaike, que lhe deu o nome de "Akaike Information Criterion" (AIC) sendo apenas apresentado em 1974. Este critério é utilizado como medida de avaliação do ajustamento de um modelo estatístico estimado, podendo ser utilizado para descrever

a relação entre a variância e o viés de construção de um modelo, demonstrando a precisão e complexidade do mesmo.

O AIC é um teste efectuado entre modelos, não sendo um teste de hipóteses a um único modelo, testa qual o melhor entre um conjunto de modelos. Dado um conjunto de dados, e vários modelos que utilizem esse mesmo conjunto de dados, o AIC classifica-os, sendo o melhor modelo o que apresentar menor AIC.

Este critério de selecção tem como base a função *Log-verosimilhança*, com a introdução de um factor de correcção de modo diminuir a complexidade do modelo.

Segundo Turkman (2000), a estatística correspondente para o modelo em H_0 é

$$AIC = -2\ell(\tilde{\beta}_1, 0, \tilde{\varnothing}) + 2r \quad (4.5.1)$$

onde $r = \dim(\beta_1)$.

A relação existente entre o AIC e o *desvio reduzido* relativo ao modelo especificado por H_0 (supõe-se que o parâmetro \varnothing é conhecido e substituído por uma estimativa consistente, neste estudo será considerado $\varnothing = 1$) é representada por:

$$\begin{aligned} AIC_r &= -2\ell(\tilde{\beta}_1, 0) + 2\ell(\hat{\beta}_S) - 2\ell(\hat{\beta}_S) + 2r \\ &= D_r^* + 2r - 2\left(\ell(\hat{\beta}_S)\right), \end{aligned} \quad (4.5.2)$$

onde r corresponde ao índice que especifica o modelo em avaliação e S refere-se ao modelo completo.

É ainda sugerida uma outra forma do modelo Akaike para seleccionar modelos. Seja,

$$C_r^* = D_r^* + 2r - n = AIC + 2\ell(\hat{\beta}_S) - n \quad (4.5.3)$$

Ao desenhar o gráfico de C_r^* sobre r obtém-se uma boa forma para comparação de modelos. Se o modelo for adequado, espera-se que C_r^* seja próximo de r .

No caso de modelos encaixados M1 e M2 com, r_1 e r_2 os respectivos parâmetros e $r_1 > r_2$ tem-se,

$$AIC_{r_1} - AIC_{r_2} = C_{r_1}^* - C_{r_2}^* = D_{r_1}^* - D_{r_2}^* + 2(r_1 - r_2)$$

e supondo que o modelo M2 é verdadeiro, tem-se

$$\mathbb{E} = [AIC_{r_1} - AIC_{r_2}] = r_1 - r_2 + 0(n^{-1})$$

Na comparação dos modelos sucessivamente mais adequados, o declive esperado do segmento de recta que une $AIC_{r_1} - AIC_{r_2}$ deve ser próximo de 1. Dois

modelos que apresentem declive maior que 1 indicam que o modelo menor é significativamente melhor que o modelo maior.

O AIC classifica um modelo pela proximidade dos seus valores com os verdadeiros valores, em termos de um determinado valor esperado. Mas torna-se importante realçar que o valor de AIC atribuído a um modelo serve apenas para classificar os modelos concorrentes, expressando qual é o melhor entre as alternativas testadas.

A metodologia AIC tem como objectivo encontrar o modelo que melhor explique os dados com um mínimo de parâmetros livres, sendo o melhor modelo, entre os modelos em comparação, o que menor AIC apresentar.

4.5.2. Avaliação dos modelos

4.5.2.1. Função Desvio

Após ser testada a significância dos parâmetros que constituem o modelo, é necessário testar a qualidade do modelo como um todo.

Segundo Turkman e Silva (2000), quando se pretende avaliar a qualidade de ajustamento de um determinado modelo de investigação (M) recorre-se ao modelo completo ou saturado (S), introduzindo-se uma medida de distância dos valores ajustados $\hat{\mu}$ com o modelo em estudo e dos correspondentes valores observados y . A distância entre o modelo completo e o modelo com menos parâmetros é medida com recurso à estatística de razão de verosimilhança de Wilks.

O logaritmo da função de verosimilhança (*função log-verosimilhança*) de um modelo linear generalizado é dado por:

$$\ln L(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{\omega_i [y_i q(\mu_i) - b(q(\mu_i))]}{\phi} + c(y_i, \phi, \omega_i) \quad (4.5.4)$$

onde $q(\mu_i) = \theta_i$ é função de ligação canónica e ω_i é o peso conhecido associado à i -ésima observação.

Ao comparar o modelo de investigação (M) com o modelo saturado (S) utilizando a estatística de razão de verosimilhança obtém-se o *desvio reduzido*:

$$\begin{aligned} D * (\mathbf{y}; \hat{\boldsymbol{\mu}}) &= -2[l_M(\hat{\boldsymbol{\beta}}_M) - l_S(\hat{\boldsymbol{\beta}}_S)] \\ &= -2 \sum_i \frac{\omega_i}{\phi} ([y_i q(\hat{\mu}_i) - b(q(\hat{\mu}_i))] - [y_i q(\mu_i) - b(q(\mu_i))]) \\ &= \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{\phi} \end{aligned} \quad (4.5.5)$$

onde $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ corresponde ao *Desvio* para o modelo em análise, sendo este apenas função do dados.

É possível definir a função desvio como,

$$\begin{aligned} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) &= \sum_i 2\omega_i \{y_i(q(y_i) - q(\hat{\mu}_i)) - b(q(y_i)) + b(q(\hat{\mu}_i))\} \\ &= \sum_i d_i \end{aligned} \quad (4.5.6)$$

correspondendo à soma de parcelas d_i , que medem a diferença dos logaritmos das verosimilhanças observada e ajustada para cada observação.

É perceptível que o *desvio* é sempre não negativo e crescente à medida que se vão retirando covariáveis ao modelo completo. Outra propriedade do *Desvio* é a aditividade para modelos encaixados. Dados dois modelos intermédios M1 e M2, com M2 encaixado em M1, ou seja, pertencem ao mesmo tipo de modelos sendo o M2 composto por menos parâmetros que o M1, designando $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ o desvio do modelo $M_j, j = 1, 2$, deste modo a estatística da razão de verosimilhança para comparar estes dois modelos é dada por:

$$-2(l_{M2}(\boldsymbol{\beta}_2) - l_{M1}(\boldsymbol{\beta}_1)) = \frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}}_2) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}_1)}{\emptyset}$$

Sob a hipótese do modelo M1 ser verdadeiro, tem-se que

$$\frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}}_2) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}_1)}{\emptyset} \sim \chi_{p_1 - p_2}^2$$

onde p_j corresponde à dimensão do vector $\boldsymbol{\beta}$ para o modelo M_j . Assim, a diferença dos *Desvios* de cada modelo pode ser utilizada como base para a comparação de modelos encaixados.

No caso de uma distribuição Gama a função desvio é dada por:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[-\log \left(\frac{y_i}{\hat{\mu}_i} \right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right] \quad (4.5.7)$$

4.5.2.2. Análise dos Resíduos

A análise de resíduos, para além de permitir avaliar a qualidade do ajustamento, pode ajudar a identificar observações que não sejam bem explicadas pelo modelo. O estudo dos resíduos permite ainda verificar os pressupostos assumidos na formulação do modelo relacionados com as variáveis explicativas, na escolha da distribuição e da função de ligação.

Uma das razões referida por muitos outros autores, para a falta de ajuste de alguns modelos está relacionada com a existência de *outliers* nos dados, no entanto é importante perceber se se trata realmente de um *outlier* ou se é o modelo que se encontra mal especificado.

Um resíduo R_i deve exprimir a discrepância entre o valor observado y_i e o valor \hat{y}_i ajustado pelo modelo. Na análise dos resíduos é possível identificar vários tipos de resíduos dos quais se destacam os resíduos de Pearson e os desvios residuais.

Considerando os dados $(Y_i, m_i, \pi_i) : 1 \leq i \leq n$, com Y_i independentemente distribuído, seguindo uma distribuição Binomial (m_i, π_i) , π_i satisfaz $\text{logit}(\pi_i) = \log\left[\frac{\pi_i}{1-\pi_i}\right] = \mathbf{x}_i^T \boldsymbol{\beta}$, com $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ik})^T$ vectores covariável conhecidos e $\boldsymbol{\beta} \in \mathbb{R}^k$ desconhecido. A função de resíduos de Pearson, escolha mais comum, pode escrever-se da seguinte forma:

$$R_i^P(y_i, \pi_i) = \frac{(y_i - m_i \pi_i)}{[m_i \pi_i (1 - m_i \pi_i)]^{1/2}}$$

Porém alguns autores sugerem a utilização do desvio residual como medida do desvio de cada observação e pode ser utilizado para verificar, nos MLG, a adequação do modelo ajustado para cada observação. O desvio residual pode ser determinado utilizando a expressão:

$$R_i^D(y_i, \pi_i) = \delta_i \left[2 \left(y_i \ln \left(\frac{y_i}{m_i \pi_i} \right) + (m_i - y_i) \ln \left(\frac{m_i - y_i}{m_i - m_i \pi_i} \right) \right) \right]^{\frac{1}{2}},$$

onde $\delta_i = \text{sin}(\pi_i - y_i)$.

Ao analisar os resíduos, se uma observação i apresentar um valor elevado de R_i^D , a observação em causa tem um contributo excessivo para o desvio apresentado, denotando-se uma má especificação do modelo para essa observação.

Os resíduos são, por norma, analisados através de gráficos dos seus valores versus os valores ajustados, permitindo assim uma melhor percepção visual da discrepância existente entre os valores observados e os estimados através do modelo.

Para analisar um gráfico que descreva os resíduos é necessário conhecer alguns conceitos como é o caso da alavancagem, influência e distância do Cook.

Alavancagem

Estatisticamente, a alavancagem ou *leverage* está muito relacionada com os modelos de regressão onde se pretende analisar observação que influenciam a análise dos resultados desses modelos. Os pontos de alavancagem são observação que se

distanciam muito da maioria dos valores das variáveis independentes, estando a uma distância considerável do valor médio de Y . Os pontos de alavancagem fazem com que o modelo, ao passar próximo de si, se afaste das restantes observações.

As medidas de alavancagem de um determinado ponto x são proporcionais a $(y - \bar{y})^2$, sendo a medida de alavancagem mais comum:

$$h_i = \frac{1}{n} + \frac{(y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

com h_i o i -ésimo elemento da diagonal da matriz H :

$$\mathbf{D}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^{1/2}$$

Onde \mathbf{D} representa a matriz diagonal $n \times n$ com $m_i \pi_i (1 - \pi_i)$ sendo o i -ésimo elemento diagonal e \mathbf{X} a matriz $n \times k$ de dimensão k com x'_i a i -ésima linha.

Influência

Pode dizer-se que um ponto é muito influente se:

$$h_i > \frac{2p}{n}$$

onde p corresponde ao número de parâmetros no modelo e n o número de observações.

No entanto a influência que os dados apresentam sobre o modelo não deve ser apenas analisada com base nos resíduos, pois um ponto com grande influência pode aproximar a linha de regressão a si, diminuindo deste modo o valor do resíduo associado. Caso, durante a análise dos dados, fossem retiradas as observações influentes seria notório o seu impacto sobre o conjunto dos valores ajustados de Y e dos parâmetros estimados $\hat{\beta}$.

Distância de Cook

De modo a analisar o grau de influência de uma observação i , é muitas vezes utilizada a distância de Cook:

$$C_i = |r_i^*| \left(\frac{n-p}{p} \cdot \frac{h_i}{1-h_i} \right)^{1/2}$$

Esta medida permite ter a percepção de qual o impacto de se retirar da análise uma observação i . Se o valor de C_i for baixo a presença dessa observação não irá

alterar muito a estimativa dos parâmetro. No entanto, caso o seu valor seja elevado a observação em causa deve ser retirada e as estimativas devem ser recalculadas, uma vez que os pontos que apresentam resíduos elevados (outliers) e/ou com alta alavancagem podem enviesar a precisão de uma regressão.

5. MODELAÇÃO DA RENTABILIDADE DE CLIENTES

Ao elaborar um estudo sobre rentabilidade de clientes é fundamental definir o conceito de rentabilidade, mas também o que define um cliente rentável e um cliente não rentável.

Dado que grande parte dos custos e proveitos de uma seguradora automóvel advêm dos custos associados aos sinistros dos seus clientes e às receitas dos prémios de seguro recebidos, neste projecto considera-se rentável um cliente cujos custos com sinistros não ultrapasse o valor pago em prémios, originando lucros para a empresa. Por sua vez, um cliente não rentável é um cliente que foi responsável por um ou mais sinistros, originando custos superiores ao valor do prémio pago à seguradora. No entanto, é importante referir que dentro dos clientes não rentáveis existem os que apresentam um nível de risco baixo, em que os sinistros são menos graves, acarretando custos não muito elevados para a companhia e os que apresentam um alto risco, quando o custo resultante do sinistro é bastante elevado.

Os dados estatísticos referentes à informação exposta serão apresentados posteriormente.

5.1. CARTEIRA DE CLIENTES

O conjunto de dados utilizado para efectuar este estudo é composto pela carteira de clientes de seguro automóvel da Seguros LOGO entre 2008 (início da actividade da empresa) e final de 2012, perfazendo aproximadamente 200.000 observações. Para a análise da rentabilidade serão contabilizados os contratos cujos prémios tenham sido pagos e os sinistros, abertos durante esse período, da responsabilidade dos segurados (e, portanto, da LOGO). Os sinistros de responsabilidade de terceiros não serão contabilizados para o presente estudo, e nos sinistros em que a responsabilidade é assumida por ambas as partes (segurado e terceiro) apenas são contabilizados os custos que ficam a cargo do segurado.

É importante realçar que a carteira em estudo é relativamente pequena, uma vez que a companhia iniciou a sua actividade em 2008, registando apenas cinco anos de actividade quando os dados foram retirados. Mesmo que o crescimento da carteira tenha sido significativo, o reduzido número de observações conduzir ao enviesamento dos resultados.

Numa primeira fase será analisada a produção, ou não, de lucro para a empresa, independentemente dos valores de lucros ou perdas, de modo a distinguir os clientes

lucrativos para a empresa. Na segunda fase, o estudo terá em conta os benefícios e custos dos segurados, sendo analisada a rentabilidade gerada por cada cliente, que resulta da diferença entre o prémio pago e os custos com sinistros, tentando identificar quais os clientes que originam grandes despesas para a seguradora.

5.1.1. Tratamento de Dados

Como referido anteriormente, os dados utilizados para a elaboração deste relatório foram recolhidos da base de dados de clientes da Seguros LOGO. Após os dados serem recolhidos e antes de se proceder ao tratamento estatístico dos mesmos, foi necessário algum tratamento de dados de modo a retirar dados inconsistentes ou repetições, que pudessem enviesar os resultados. Para além da “limpeza” dos dados foram criadas novas variáveis resultantes da conjugação de outras, a fim de se obter a informação que foi considerada necessária e relevante para a análise da rentabilidade dos clientes.

De acordo com o que é usual fazer-se na análise de risco em tarificação, e utilizando como base a informação cedida pela seguradora, algumas das variáveis foram agrupadas em classes de risco, como por exemplo, a idade, o tempo de carta ou de seguro. Por motivos de sigilo, os intervalos utilizados não serão apresentados. A sintetização da informação permite uma modelação e posterior interpretação de resultados mais fácil e clara.

Após a elaboração dos passos referidos, o universo sob o qual irá incidir a restante análise é composto por 190.127 observações.

5.1.2. Variáveis Utilizadas

Para avaliar a rentabilidade dos clientes da Seguros LOGO foram criados dois modelos que permitem identificar as características que possam definir o perfil de um cliente rentável.

Dado que muita da rentabilidade de um cliente de uma seguradora automóvel é explicada pelos custos que a seguradora poderá ter durante a vigência dos contratos dos clientes, optou-se por utilizar como variáveis resposta para os modelos criados a *Produção de Lucro* (produção ou não de lucro), e a *Rentabilidade Gerada* (montante que relaciona o prémio recebido com os custos de sinistros).

Relativamente às variáveis independentes foram seleccionadas de acordo com os dados que são recolhidos nas simulações e que são utilizados na tarificação por grande

parte das empresas de seguros, variáveis estas escolhidas de acordo com estudos feitos pelas próprias empresas ou por entidades externas.

No modelo onde se pretende avaliar a geração de lucros, a variável dependente é uma variável *dummy*, podendo apenas assumir dois valores, caso o cliente tenha gerado lucros, a variável assume o valor um (1) em caso contrário assume o valor zero (0). No modelo da rentabilidade gerada a variável resposta é numérica assumindo os valores reais das variáveis (montante resultante da diferença entre o valor pago pelo cliente e o valor pago pela seguradora, no caso da ocorrência de sinistros).

Para a construção dos dois modelos foram consideradas as mesmas variáveis independentes, presentes na Tabela 5.1, sendo estas variáveis maioritariamente categóricas (variáveis qualitativas, que expressam uma característica que pode ou não assumir uma ordem), sendo também consideradas algumas variáveis numéricas.

VARIAVEL	CATEGORIA
DIAS_CONTRATO	(VARIÁVEL DISCRETA)
CANAL	CALL_CENTER, INTERNET, PARCEIROS
PRODUTO	LIGHT, TOPPIN, SAFE, MAX, MOTO, LEASE
FRACCIONAMENTO	MENSAL, TRIMESTRAL, SEMESTRAL, ANUAL
BONUS_MALUS	(VARIÁVEL DISCRETA)
TEMPO_SEGURO	TS1; TS2; TS3; TS4;TS5
IDADE_VEICULO	IV1; IV2; IV3; IV4
TIPO_VEICULO	B, C, L, M
CILINDRADA	C1 a C5
COMBUSTIVEL	D, E, G, M
LUGARES	L1 a L9
PREMIO_TOTAL	(VARIÁVEL CONTINUA)
SEXO	F, M
IDADE	I1 a I8
ANOS_CARTA	TC1 a TC5
ZONA	Z1 a Z9
QIV	N;S
VS	N;S
FRANQUIA	F0, F2, F4
AV_FN	N;S
IE	N;S
SIN_5_ANOS	SIN1 a SIN4

Tabela 5.1 - Variáveis Independentes

Descrição das variáveis na secção ANEXOS (9.1.1.4 - **Variáveis Utilizadas**).

5.2. PRODUÇÃO DE LUCRO

Variável dependente:

Produção de Lucros (v.a. R): Variável *dummy* que assume o valor 1 no caso de o cliente ter originado lucro para a empresa e o valor 0 em caso contrário.

Variáveis Independente:

Variáveis indicadas na Tabela 5.1 - Variáveis Independentes.

5.2.1. Produção de Lucro - Análise Estatística das variáveis

Recorrendo ao *software* R Project¹ foi efectuada uma análise estatística dos dados da base de dados da Seguros LOGO. Os primeiros resultados obtidos a partir do R foram as estatísticas descritivas tendo por base os dados que caracterizam a carteira da seguradora.

Variáveis Numéricas (Produção de lucro):

Na Tabela 5.2 encontra-se a percentagem de contratos que não produziu lucro e os que produziram lucros para a seguradora. Verifica-se que existe uma pequena percentagem de clientes que não produziu lucros para a empresa, contudo analisar apenas a proporção de clientes nessas circunstâncias pode conduzir a conclusões erradas. Como tal, procedeu-se também à análise da rentabilidade gerada pelos clientes, tendo como base a diferença entre os prémios pagos e o custo dos sinistros (análise mais detalhada da rentabilidade gerada no capítulo 0).

	R (Produção de Lucros)
Não Produziu (0)	9%
Produziu (1)	91%
	Rentabilidade Gerada
Máximo	5.648,37 €
Mínimo	- 44.400,89 €
Média	128,50 €

Tabela 5.2 - Produção de Lucro e Rentabilidade Gerada

Analisando sem grande detalhe a carteira de clientes verifica-se que aproximadamente 91% dos contratos produziu lucros para a empresa, enquanto apenas 9% não produziu. Porém, o que à partida não aparenta ser um cenário muito

¹ The R Project for Statistical Computing (<http://www.r-project.org/>)

pessimista pode não reflectir os reais encargos da seguradora. Note-se que os prémios de seguro, calculados como a esperança matemática do risco, à qual se adicionam cargas de segurança, encargos, impostos e taxas, não são valores extremamente elevados (têm por base a lei dos grandes números e o princípio de mutualização entre apólices) mas que, por outro lado, os montantes a pagar por indemnizações decorrentes de sinistros automóveis podem ter valores muito avultados.

Na tabela que se segue é possível observar na primeira coluna de cada subtabela a média de cada variável numérica, na segunda coluna a diferença percentual que se verifica entre o valor máximo, mínimo e médio quando se compara o total de contratos com o total de contratos que não produziram lucros e na última coluna encontram-se as mesmas métricas mas comparativamente ao total de contratos que produziram lucro.

Ao analisar a variável DIAS_CONTRATO verifica-se que, comparativamente à totalidade de cliente (TOTAL), a permanência na companhia é, em média, 9% superior nos clientes que não produziram lucro (N_PROD_L) para a empresa. Por sua vez a permanência dos que produziram lucro (PROD_L) é, em média, 1% inferior.

Por razões de confidencialidade não serão apresentados os valores absolutos.

	DIAS_CONTRATO			BONUS_MALUS		
	TOTAL	N_PROD_L	PROD_L	TOTAL	N_PROD_L	PROD_L
Máximo		0%	0%		0%	0%
Mínimo		0%	0%		0%	0%
Média	581	+9%	-1%	57	0%	-2%
	PREMIO_TOTAL					
	TOTAL	N_PROD_L	PROD_L			
Máximo		-92%	+48%			
Mínimo		+23%	-29%			
Média	219	+9%	-12%			

Tabela 5.3 - Análise Preliminar - Carteira (variáveis numéricas)

Na **Erro! A origem da referência não foi encontrada.** é possível observar que ace ao total de contratos, os clientes que não produziram lucro (N_PROD_L) permanecem, em média, mais tempo na companhia que a generalidade dos clientes, o que pode não ser muito vantajoso para a empresa, que pode estar a perder clientes com rentabilidade mais elevada. Note-se que o tempo médio de permanência em carteira, dos clientes que geraram lucro é 1% inferior à média da carteira enquanto que, relativamente aos que não produziram lucro, o seu tempo de permanência é, em média, 9% superior à média da carteira.

Relativamente ao nível de Bónus-Malus que lhes é atribuído quando o contrato é tarifado verifica-se que os clientes que produzem lucros apresentam, em média, um valor ligeiramente mais baixo. O facto de apresentarem à partida um nível de risco inferior reflecte-se no prémio que lhes é conferido, sendo em média inferior ao prémio médio do total de clientes.

Variáveis Categóricas (Produção de lucro):

Ao observar a Tabela 9.2 – Produção de Lucros - Análise Preliminar), presente nos Anexos, é perceptível que o canal de distribuição que capta mais clientes é o dos Parceiros, sendo o peso mais significativo no conjunto de clientes que não produz lucro. O produto mais comercializado é o Logo Topping sendo a preferência de mais de 50% do total de clientes, percentagem também verificada quando se desagrupa os clientes de acordo com a sua produção de lucros. É de realçar que a percentagem de clientes que opta pelo Logo Max é superior nos clientes que não produzem lucro comparativamente ao que produzem, reflexo dos custos que a empresa tem de suportar com a componente de danos próprios. Quando se analisa o Logo Moto o cenário é o oposto, dentro dos clientes sem lucros gerados apenas 4% optaram por este produto contra os 12% da carteira. O fraccionamento preferencial é o anual com 39% da carteira, sendo esta percentagem bastante mais pequena quando se olha para o universo dos clientes não lucrativos (34%), passando o pagamento trimestral a estar em primeiro lugar com 36%. Quanto às coberturas apenas 20% subscreve a cobertura de Veículo de Substituição (VS) e 9% Indemnização Extra (IE), 7% subscreve Actos de Vandalismo e Fenómenos da Natureza (AV_FN), sendo a Quebra Isolada de Vidros (QIV) a cobertura com maior percentagem de subscritores com 24% no total da carteira. Novamente estas percentagens são diferentes quando restringe a observação aos clientes com custos superiores ao proveito, 24% subscreve VS, 10% AV_FN e 25% QIV. Quanto à franquia, uma larga maioria (95% da carteira total) opta por deixar a cargo da seguradora o total das despesas com os possíveis sinistros.

Os clientes que mais procuram a LOGO são do sexo masculino, representando 75% da procura, sendo essa percentagem menor nos clientes que não originam lucro (71%). Os clientes da seguradora encontram-se, na sua maioria, entre as idades I4 e I6 (72%) e pertencem ao conjunto de clientes com mais anos de carta e de seguro (TC5 com 88% e TS5 com 75%). Os valores de tempo de carta são ligeiramente menores quando se olha apenas para os clientes com prejuízos associados, continuando no entanto estes intervalos a deter a maioria dos clientes (85% nos anos de carta e 76%

no tempo de seguro). A grande maioria dos clientes LOGO reside entre as zonas 2 e 6 (89%).

Os veículos que compõem a carteira da LOGO são principalmente da classe L (79% no total da carteira e 87% no total de sinistrados), com alguns anos (IV3) com 65% na carteira vs 67% nos sinistrados, a gasolina (59% carteira vs 65% nos não lucrativos) e com 5 lugares (73% na carteira vs 81% nos não lucrativos).

De acordo com os valores apresentados, comparativamente com a média da carteira, um cliente que não produz lucros permanece mais tempo na companhia, apresenta escalão de Bonus-Malus ligeiramente mais elevado, o que reflecte níveis de risco superiores, originando prémios mais elevados. Subscvem mais frequentemente o seguro por telefone, optando maioritariamente pelo Logo Topping, com fraccionamento trimestral. Os veículos são relativamente antigos (IV3), ligeiros de passageiros, com cilindradas não muito elevadas (C2), a gasolina e de 5 lugares. A experiência de carta e seguro dos segurados é elevada (TC5 e TS5), com idade média (I4, I5 e I6), do sexo masculino, subscvendo o seguro sem coberturas facultativas (QIV, VS, AV/FN e IE).

Para uma melhor visibilidade da distribuição dos dados dentro de cada variável em análise procedeu-se à representação gráfica das mesmas utilizando histogramas.

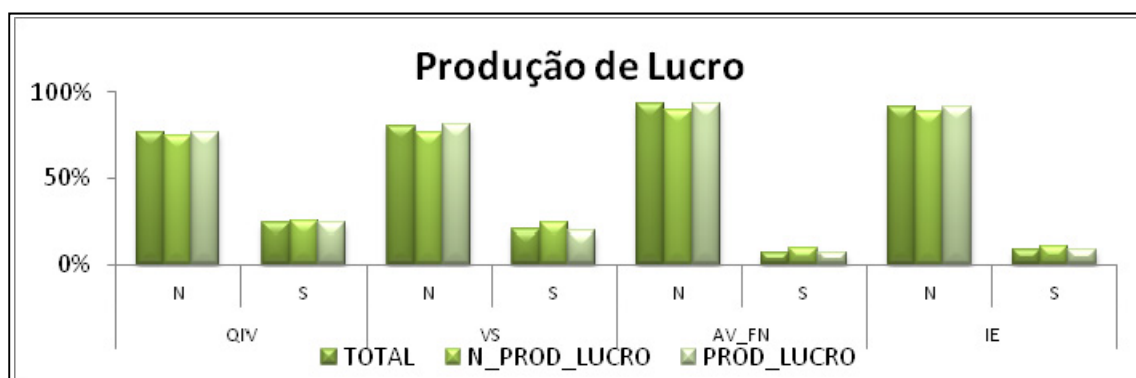


Figura 5.1 - Produção de Lucro - Coberturas

5.2.2. Produção de Lucro - Resultados

Nesta secção aplicaram-se métodos estatísticos para elaboração do estudo da carteira de clientes, dado o prémio pago e o histórico de sinistralidade.

Numa primeira análise apenas será avaliada a probabilidade de um cliente produzir lucro para a empresa.

Com o intuito de avaliar a probabilidade de um cliente gerar, ou não, lucro para a companhia será utilizada a regressão logística para ajustar os modelos e, os métodos

Stepwise-backward e AIC na escolha das variáveis que melhor explicam a variável resposta e que constituem o modelo final.

De acordo com a forma dos modelos lineares,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon, i = 1, \dots, n \quad (5.2.1)$$

onde β_i corresponde ao coeficiente discriminante para a variável x_i e ε um vector de erros aleatórios, o modelo criado para analisar a probabilidade de um cliente produzir lucro para a empresa, tem como variável resposta a produção ou não de lucros (R), que assume o valor 1 em caso de produzir lucros e o valor 0 em caso contrário. O conjunto de variáveis independentes é formado pelas variáveis: Dias de contrato, Canal, Produto, Fraccionamento, Bónus-Malus, Tempo de seguro, Idade do veículo, Tipo de veículo, Cilindrada, Combustível, Lugares, Prémio total, Sexo, Idade, Anos de carta, Zona, Quebra Isolada de Vidros, Veículo de Substituição, Franquia, Actos de Vandalismo e Fenómenos da Natureza, Indemnização Extra e Número de Sinistros nos últimos 5 anos.

Utilizando o *Software R* é possível ajustar um MLG a partir da função *glm*, onde se especificam as variáveis preditivas do modelo e a respectiva distribuição dos erros. A função de ligação que será utilizada é a função logit, que se omitirá nos ajustamentos seguintes uma vez que é a função utilizada por defeito pelo *Software R* para o modelo com erros binomiais.

(O estudo que se segue foi efectuado com recurso ao Software R.)

O modelo completo contém, portanto, as seguintes covariáveis:

Dias de contrato, Canal, Produto, Fraccionamento, Bónus-Malus, Tempo de seguro, Idade do veículo, Tipo de veículo, Cilindrada, Combustível, Lugares, Prémio total, Sexo, Idade, Anos de carta, Zona, Quebra Isolada de Vidros, Veículo de Substituição, Franquia, Actos de Vandalismo e Fenómenos da Natureza, Indemnização Extra e Número de Sinistros nos últimos 5 anos.

Modelo Nulo

Iniciou-se a modelação ajustando o modelo nulo (com as covariáveis iguais a zero) obtendo-se os seguintes resultados:

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1616	0.4509	0.4509	0.4509	0.4509

AIC: 120817

Modelo Completo

Realizou-se a mesma análise para o modelo completo (contendo todas as covariáveis)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9835	0.3377	0.4223	0.4803	1.4492

AIC: 115854

Os restantes outputs do Software R encontram-se na secção Anexos (9.2.3.1 - Produção de Lucro - Resultados)

Comparação entre os dois modelos

Após a obtenção dos resultados dos dois modelos realizou-se a comparação dos mesmos através do *Teste de Razão de Verossimilhanças*, testando se todas as covariáveis são iguais a zero. A não rejeição dessa hipótese indicaria que o modelo sem as covariáveis seria significativamente melhor que o modelo com todas as covariáveis em análise.

As hipóteses testadas são:

$$H_0: (\hat{\beta}_{DIAS_CONTRATOS}, \dots, \hat{\beta}_{SIN5_ANOS}) = \mathbf{0}$$

vs

$$H_1: (\hat{\beta}_{DIAS_CONTRATOS}, \dots, \hat{\beta}_{SIN5_ANOS}) \neq \mathbf{0}$$

Os resultados obtidos são:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190126	120815			
2	190059	115718	67	5096.4	< 2.2e-16

De acordo com os resultados obtidos, com base no *Test de Wilks*, é possível verificar que o *p-value* é bastante pequeno (menor que 2×10^{-16}) e, como tal rejeita-se a hipótese do modelo nulo ser estatisticamente mais significativo que o modelo completo. Por outras palavras, rejeita-se a hipótese de os coeficientes acrescidos serem iguais a zero, existindo pelo menos um com significância estatística. Para além dos resultados do teste de *Wilks*, também se podem comparar os valores do Critério de Informação de Akaike (AIC) dos dois modelos. O modelo nulo apresenta um AIC de

120817 enquanto o modelo completo apresenta um AIC de 115854, o que mais uma vez indica que o modelo contendo todas as covariáveis é melhor que o modelo nulo, pois apresenta um menor AIC.

Seleção do modelo através do Método Stepwise-Backward

O modelo completo, contendo todas as covariáveis, apresentado no tópico anterior servirá com modelo inicial para a modelação através do método Stepwise-Backward. Este método consiste em retirar as covariáveis que sejam menos significativas, de acordo com o Teste de Wald.

Ao observar as estimativas dos parâmetros para o modelo completo, presentes na secção Anexos (9.2.3.1 - Produção de Lucro - Resultados), é perceptível que existem várias covariáveis com significância estatística, para um nível de confiança de 95% (valores de p-value inferior a 0,05), porém, também existem covariáveis com valores de p-values elevados, apresentando baixa significância estatística. Considerando tais valores, e segundo o método *Stepwise*, a primeira covariável a ser retirada é a que apresentar maior *p-value*, que neste caso é a covariável Idade (IDADE). Após esta ser retirada estimam-se os parâmetros do novo modelo.

A fim de se validar a hipótese de o novo modelo ser estatisticamente mais significativo que o anterior, procede-se à comparação entre ambos através do Teste de Razão de Verosimilhança:

$$H_0: \hat{\beta}_{IDADE} = \mathbf{0} \quad \text{vs} \quad H_1: \hat{\beta}_{IDADE} \neq \mathbf{0}$$

para o qual se obteve a seguinte informação:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190066	115752			
2	190059	115718	7	34.021	1.707e-05

Analisando o p-value do teste conclui-se que a covariável Idade deve ser mantida no modelo. Comparando o Modelo Completo e o modelo sem a covariável, com base no Teste de Wilks, rejeita-se a hipótese que defende que os parâmetros $\hat{\beta}_{IDADE}$ são iguais a zero (p-value < 0,05), mantendo-se o Modelo Completo como o melhor ajustamento até ao momento.

Prossegue-se a análise, considerando o modelo com a covariável Idade e retirando-se a próxima covariável com o p-value mais elevado, sendo esta covariável Actos de Vandalismo e Fenómenos da Natureza (AV_FN).

Retira-se a covariável, volta a estimar-se o modelo e procede-se novamente à comparação com o Modelo Completo (melhor modelo da etapa anterior):

$$H_0: \hat{\beta}_{AV_FN} = \mathbf{0} \quad \text{vs} \quad H_1: \hat{\beta}_{AV_FN} \neq \mathbf{0}$$

Aplicando o Teste de Wilks, obtém-se:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190060	115718			
2	190059	115718	1	0.028004	0.8671

Observando o p-value, conclui-se que o modelo sem a covariável Actos de Vandalismo e Fenómenos da Natureza é estatisticamente mais significativo que o Modelo Completo ($p\text{-value} > 0,05$).

Proseguiu-se a análise, avaliando a exclusão ou não das restantes covariáveis que apresentam menor significância estatística.

Ao observar os resultados dos Testes de Razão de Verossimilhança que comparam os modelos sem as covariáveis com o modelo resultante da etapa anterior verifica-se que existem covariáveis com pouca significância estatística e que não acrescentam grande valor ao modelo, podendo desta forma ser excluídas do modelo final. Após ser testada a exclusão das covariáveis que apresentam pouca significância estatística, foram retiradas as covariáveis Actos de Vandalismo e Fenómenos da Natureza (AV_FN), Indemnização Extra (IE), Quebra Isolada de Vidros (QIV) e Bónus-Malus (BM), obtendo-se o modelo composto pelo conjunto de variáveis que melhor explicam a probabilidade de produção de lucro (Modelo Stepwise).

Para uma análise mais detalhada, o restante código e resultados encontram-se nos Anexos (9.2.3.3 - Selecção do modelo através do Método Stepwise-Backward).

Selecção das variáveis através do Critério de Informação de Akaike (AIC)

Outro método de selecção de modelos é através do Critério de Informação de Akaike, já descrito anteriormente no Capítulo 4.5.1.3. - Método AIC.

O modelo obtido através desta metodologia estima que o modelo estatisticamente mais adequado para modelar o fenómeno em estudo deve conter as covariáveis: Dias de contrato, Canal, Produto, Fraccionamento, Bónus-Malus, Tempo de seguro, Idade do veículo, Tipo de veículo, Cilindrada, Combustível, Lugares, Prémio total, Sexo, Idade, Anos de carta, Zona, Veículo de Substituição, Franquia e Número de Sinistros nos últimos 5 anos.

Face ao melhor modelo resultante da aplicação do método Stepwise-Backward (Modelo Stepwise), a diferença no modelo obtido (Modelo AIC) reside apenas na não exclusão da covariável Bónus-Malus, sendo o valor de AIC o mesmo em ambos os modelos (AIC=115849).

Aplicando o Teste de Wilks, para comparar o modelo obtido através da metodologia AIC com o melhor modelo obtido através do método Stepwise-Backward, obtém-se:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190062	115719			
2	190063	115721	-1	-2.2899	0.1302

Ao analisar o p-value obtido verifica-se que através do método Stepwise-Backward o conjunto de covariáveis que compõem o modelo é estatisticamente mais significativo do que o conjunto obtido pela metodologia do Critério de Informação de Akaike.

Observando o Modelo Stepwise é visível que existem determinadas categorias dentro das covariáveis que não são significativamente diferentes da categoria definida como nível base do factor e, a fim de simplificar o modelo, optou-se por agrupar as categorias não significativas à base.

De modo a validar se o agrupamento das categorias ao nível base deve ou não ser efectuado, realizou-se o teste de Wald, com recurso ao software *R*. Os resultados foram aferidos com base no valor do p-value obtido para os testes efectuados.

Após serem efectuados todos os testes necessários, as categorias agrupadas foram:

PRODUTO: LEASE+LIGHT+SAFE+TOPPING=TOPPING
 TIPO_VEICULO: B+L =L
 CILINDRADA: C1+C2+C3+C5=C1
 COMBUSTIVEL: D+E=D
 LUGARES: L1+L2+L8=L2
 IDADE: I1+I2+I6=1; I3+I4+I5=I3; I7+I8=I7
 ANOS_CARTA: TC1+TC2+TC3+TC4=TC1
 ZONA: 1+2=1

Na Tabela 5.4 encontram-se os valores dos Critérios de Informação de Akaike dos principais modelos ajustados.

Modelo	AIC
Modelo Completo	115854
Modelo Stepwise	115849
Modelo StepAIC	115849
Modelo Agrupado	115902

Tabela 5.4 - Critérios de Informação de Akaike

De acordo com os dados da tabela acima, verifica-se que os modelos Stepwise e StepAIC apresentam melhores valores de AIC, não apresentando grandes diferenças face ao Modelo Completo. O Modelo Agrupado apresenta o valor de AIC mais elevado, no entanto será o modelo utilizado para a restante análise da probabilidade de produção de lucros dado ser o modelo menos complexo, permitindo análises mais simples e resultados mais claros.

5.2.3. Produção de Lucro - Análise dos Resíduos

De acordo com descrito anteriormente, a análise dos resíduos permite avaliar a qualidade do ajustamento do modelo, permitindo também perceber a existência de possíveis *outliers* que poderão enviesar a estimação.

Uma das formas de analisar os resíduos é através da sua representação gráfica, como se pode observar nas figuras que se seguem.

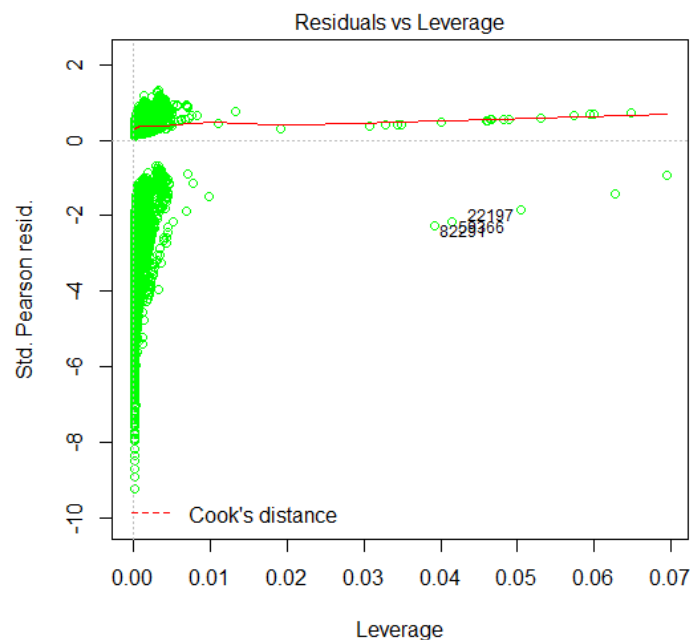


Figura 5.2 - Resíduos de Pearson Padronizados vs Alavancagem

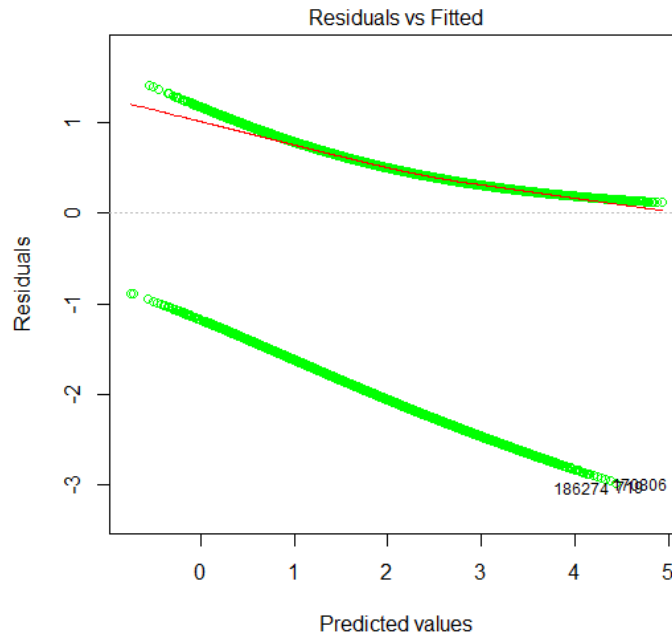


Figura 5.3 - Desvios Residuais vs Valores Ajustados

Ao observar a Figura 5.2 é perceptível que existem vários pontos cuja alavancagem é elevada distanciando-se muito dos restantes valores das variáveis independentes, podendo levar a um afastamento do valor médio da variável dependente (Produção de Lucro). Porém, todos estes pontos apresentam uma baixa Distância de Cook, não sendo pontos com demasiada influência sobre a estimação do modelo, não afastando a linha de regressão da globalidade dos pontos.

O desvio residual representa a medida do desvio do contributo de cada observação para a explicação da variável dependente, permitindo analisar a adequação do modelo ajustado a cada observação. Os valores dos resíduos, para cada categoria da variável resposta, devem ser iguais em magnitude, mas com sinal oposto. Na Figura 5.3 é possível observar que na maior parte dos casos estas condições são respeitadas, encontrando-se o modelo ajustado adequado à maioria das observações.

5.3. ESTIMAÇÃO DA PROBABILIDADE DE PRODUZIR LUCROS

Após o estudo estatístico dos dados e aplicação de métodos para selecção das covariáveis definiu-se o modelo final que servirá de base à estimação da probabilidade de um cliente, que reúna determinadas características, produzir ou não lucro à seguradora. O modelo final contém as seguintes covariáveis independentes: Dias de Contrato, Canal, Produto, Fracionamento, Tempo de Seguro, Idade do Veículo, Tipo de Veículo, Cilindrada, Combustível, Lugares, Prémio Total, Sexo, Idade, Anos de Carta, Zona, Veículo de Substituição, Franquia e Sinistros nos últimos 5 anos.

Do ajustamento do Modelo Agrupado resultam os coeficientes β_i que serão utilizados para a estimação da probabilidade de produção de lucros para a empresa. Tendo por base o modelo e os coeficientes estimados, serão apresentados alguns exemplos da estimação da probabilidade de um cliente gerar ou não lucro para a empresa.

As estimativas dos coeficientes e o modelo utilizados podem ser consultados na secção Anexos (9.1.1.7- Estimação da Probabilidade de Produzir Lucros).

É de realçar que uma categoria de cada variável categórica está associada ao parâmetro β_0 . Definiremos o conjunto de tais características como o perfil do *Cliente Padrão*, sendo, mais concretamente, definidas por:

CANAL	CALL_CENTER
PRODUTO	MAX
FRACCIONAMENTO	ANUAL
TEMPO_SEGURO	TS1
IDADE-VEICULO	IV1
TIPO_VEICULO	C
CILINDRADA	C1
COMBUSTIVEL	D
LUGARES	L2
SEXO	F
IDADE	I1
ANOS_CARTA	TC1
ZONA	Z1
VS	N
FRANQUIA	F0

SIN_5_ANOS	SIN1
------------	------

Tabela 5.5 - Características do *Cliente Padrão*

A estimação da probabilidade de produção de lucros de cada cliente, π_i , num modelo de regressão logística, é efectuada a partir de:

$$\pi_i = \frac{e_i^y}{1 + e_i^y}$$

De modo a simplificar a análise do modelo obtido, na Tabela 5.6 encontram-se representados alguns exemplos da aplicação do modelo para estimação da probabilidade de clientes com determinadas características produzirem lucros para a seguradora.

	Cliente A	Cliente B	Cliente C	Cliente D	Cliente E
DIAS_CONTRATO	600	1500	1200	300	300
CANAL	PARCEIROS	INTERNET	PARCEIROS	CALL_CENTER	CALL_CENTER
PRODUTO	MAX	MOTO	TOPPING	TOPPING	TOPPING
FRACCIONAMENTO	SEMESTRAL	ANUAL	SEMESTRAL	TRIMESTRAL	TRIMESTRAL
TEMPO_SEGURO	TS3	TS3	TS2	TS1	TS1
IDADE_VEICULO	IV3	IV4	IV3	IV1	IV1
TIPO_VEICULO	L	M	C	L	L
CILINDRADA	C1	C1	C1	C4	C4
COMBUSTIVEL	D	G	G	D	D
LUGARES	L5	L2	L2	L7	L7
PREMIO_TOTAL	350	220	100	280	280
SEXO	M	M	F	F	F
IDADE	I7	I3	I1	I1	I1
ANOS_CARTA	TC5	TC5	TC1	TC1	TC1
ZONA	Z5	Z3	Z4	Z6	Z6
VS	S	N	S	N	N
FRANQUIA	F4	F0	F0	F2	F2
SIN_5_ANOS	SIN2	SIN1	SIN4	SIN1	SIN3
$\frac{e_i^y}{1 + e_i^y}$	74,1%	96,6%	33,8%	95,8%	83,9%
IC	[69,5%;78,2%]	[95,9%;97,2%]	[27,0%;41,4%]	[95,1%;96,5%]	[81,2%;86,3%]

Tabela 5.6 - Probabilidade de produção de lucro – Exemplos

Com base nos exemplos presentes na Tabela 5.6 pode dizer-se que o Cliente B apresenta uma maior probabilidade de produzir lucro que os restantes exemplos. Pode também afirmar-se que um cliente com estas características tem associada uma probabilidade de produção de lucro de entre 96% e 97% aproximadamente, a um nível de confiança de 95%.

Por sua vez um cliente que reúna as mesmas características do Cliente C, tem uma menor probabilidade de produzir lucro, comparativamente com os restantes exemplos apresentados. Observando o intervalo de confiança verifica-se que o valor de probabilidade do limite superior não ultrapassa os 41%, ficando ainda distante dos valores dos demais casos.

Ao comparar o Cliente A com o Cliente B, onde o primeiro contratou um produto mais completo (MAX) e com um prémio de seguro mais elevado, e o segundo cliente contratou o um produto mais simples (MOTO) e com um prémio mais baixo, é perceptível que a uma maior probabilidade de produção de lucro não se baseie exclusivamente na contratação de produtos mais complexos e mais caros. Existem muitas outras características a ter em conta e que se reflectem significativamente na possível produção de lucro.

Os exemplos Cliente D e Cliente E partilham das mesmas características, com excepção do número de sinistros nos últimos 5 anos, no entanto a probabilidade de produzirem lucro é muito diferente. Um cliente que apresente determinadas características e um número de sinistros nos últimos 5 anos pertencente à classe SIN1 apresenta uma maior probabilidade de produzir lucros comparativamente a um cliente com as mesmas características, mas que nos últimos 5 anos tenham sinistros dentro da classe SIN3. A probabilidade do primeiro cenário (cliente D) situa-se entre 95% e 97%, enquanto no segundo cenário (Cliente E) a probabilidade máxima não ultrapassa os 86%, a um nível de confiança de 95%. Deste modo, pode afirma-se que o número de sinistros nos últimos 5 anos é uma característica que apresenta um grande impacto sobre a probabilidade da Produção de Lucro.

5.4. RENTABILIDADE GERADA

Variável dependente:

Rentabilidade Gerada (RENT_POSIT): Variável onde se encontra expresso o montante corresponde à diferença entre o valor pago pelo cliente, relativo ao prémio de seguro, e os custos da seguradora em caso de sinistro.

Variáveis Independentes:

Serão utilizadas as mesmas variáveis independentes anteriormente aplicadas na construção do modelo para avaliar a produção de lucros.

Variáveis indicadas na Tabela 5.1 - Variáveis Independentes.

5.4.1. Rentabilidade Gerada - Análise Estatística das variáveis

Como referido em capítulos anteriores, não é suficiente analisar apenas a produção ou não de lucro por parte de um cliente, mas sim conjugar essa análise com o proveito ou custo da empresa com os seus clientes, para que se tenha a real percepção da rentabilidade da carteira.

Variáveis Numéricas (Rentabilidade):

Como se pode observar na Tabela 5.7, em média a rentabilidade gerada é positiva, contudo apresenta um valor não muito elevado, que pode não ser suficiente para a constituição dos requisitos necessários para que a seguradora cumpra com a totalidade das suas responsabilidades e possível geração dos lucros pretendidos.

	RENTABILIDADE
Máximo	5.648,37 €
Mínimo	-44.400,11 €
Média	128,50 €
Desvio Padrão	1.105,66 €

Tabela 5.7 - Rentabilidade gerada

A análise estatística das restantes variáveis numéricas, Dias de Contrato, Bónus-Malus e Prémio total, que serão utilizadas na modelação da rentabilidade gerada, já se encontra descrita no capítulo 5.2.1 - Produção de Lucro - Análise Estatística das variáveis, dado serem as mesmas variáveis utilizadas na modelação da Produção de Lucro.

Variáveis Categóricas (Rentabilidade):

Consultando a Tabela 9.3 (disponível nos Anexos), onde se encontra a rentabilidade média por contrato, é possível verificar que os contratos subscritos no canal Parceiros apresentam um valor médio mais baixo de rentabilidade que os restantes canais.

Analisando o produto, a rentabilidade gerada é maior nos produtos de danos próprios, Logo Lease e Logo Max, o que é expectável, uma vez que os prémios cobrados são por norma mais elevados. O Logo Light e Logo Moto apresentam valores bastante mais baixos de rentabilidade, o que pode estar relacionado com o facto de terem associado um menor prémio de seguro. No caso do Logo Moto os prémios de seguro são, na sua maioria, relativamente baixos, mas o que diminui ainda mais a média da rentabilidade é a elevada ocorrência de sinistros que envolvem danos corporais, uma vez que este tipo de danos acarreta altos custos para a companhia devido a tratamentos hospitalares e indemnizações.

Relativamente às coberturas subscritas os clientes mais rentáveis são os que optam por uma protecção mais completa adicionando coberturas facultativas o que leva a um aumento do prémio de seguro, sendo Actos de Vandalismo e Fenómenos da Natureza (AV_FN) a cobertura que, quando subscrita, engloba os clientes mais rentáveis. Este facto pode ainda reflectir que os segurados que procuram uma maior protecção são indivíduos preocupados com a sua segurança o que acaba por se traduzir numa menor sinistralidade. Quanto ao fraccionamento a rentabilidade média é superior nos contratos anuais, onde a probabilidade de não pagamento da totalidade do prémio é mais reduzida.

Analisando as características do segurado é perceptível que os clientes do sexo Feminino se revelam mais rentáveis. Ao observar os valores de rentabilidade desagrupando por classes de idade (I), tempo de carta (TC) e experiência de seguro (TS), verifica-se que são mais elevados nas classes mais maduras (I8), com maior experiência de carta (TC5), mas com seguro em nome próprio há apenas alguns anos (TC3). Relativamente à zona de risco a rentabilidade gerada é, por norma, crescente da Zona1 para a Zona9.

Quando o foco é o objecto seguro os valores de rentabilidade são mais elevados nos veículos do tipo Comercial ou Ligeiro (B), o oposto verifica-se nos veículos comerciais (C) e nas motos (M). Os contratos de seguros para veículos mais recentes (IV1) estão normalmente associados à subscrição de produtos de danos próprios, com prémios mais elevados originando níveis de rentabilidade mais elevada. Quanto às

restantes características do objecto seguro as classes que se apresentam como sendo mais rentáveis são as de cilindrada C2 e C3, veículos com 7 ou 8 lugares, sendo veículos mais familiares, tendo os condutores por regra uma condução mais defensiva. Relativamente ao tipo de veículo os eléctricos têm associados valores mais altos de rentabilidade, apresentam também um valor de mercado mais elevado e não atingem velocidades muito elevadas.

Relativamente à variável Rentabilidade procedeu-se ainda à representação gráfica das variáveis utilizando histogramas e diagramas “Caixa-e-Bigodes”, permitindo uma melhor visibilidade da distribuição dos dados dentro de cada variável.

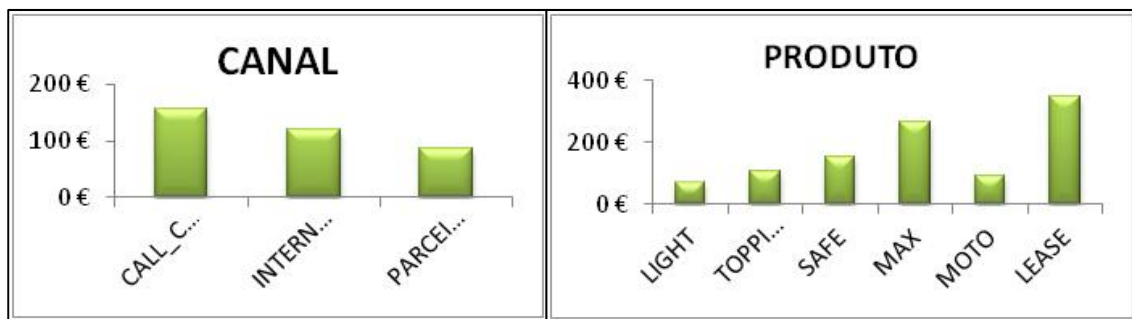


Figura 5.4 - Rentabilidade média por Produto e Canal

O diagrama “Caixa-e-Bigodes” consiste numa representação gráfica simples e de fácil construção e interpretação da variável em análise, no caso a Rentabilidade.

Neste diagrama estão representados o 1º e 3º quartis que delimitam a caixa, no interior da caixa está representada a mediana, e unidos à caixa através de pontos encontram-se o valor mínimo (ligado ao 1º quartil) e máximo (ligado ao 3º quartil) presentes no conjunto de dados analisados.

Caixa e Bigodes - Rentabilidade

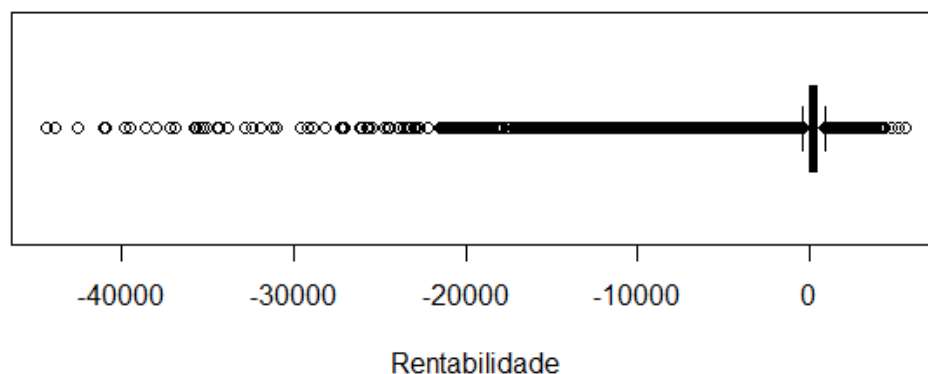


Figura 5.5 - Rentabilidade Gerada

Para a rentabilidade da carteira em análise, recorrendo ao *Software R*, obteve-se os valores correspondentes aos pontos representados no diagrama acima representado.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-44400.00	75.48	196.70	128.50	403.00	5648.00

Grande parte das observações encontra-se próximas do zero, no entanto é visível que existem alguns pontos que se afastam da generalidade das observações, estando o valor mínimo bastante afastado.

Para uma melhor percepção da rentabilidade gerada pela maioria das observações foram calculadas barreiras de outliers fazendo com que valores extremos não se encontrem entre o mínimo e o máximo representados no gráfico. Normalmente é utilizada apenas uma barreira que identifica os outliers ligeiros, mas em certos casos é necessário o cálculo de segundas barreiras devido à existência de outliers severos (valores que se afastam muito dos valores considerados observações normais).

Caixa e Bigodes - Rentabilidade

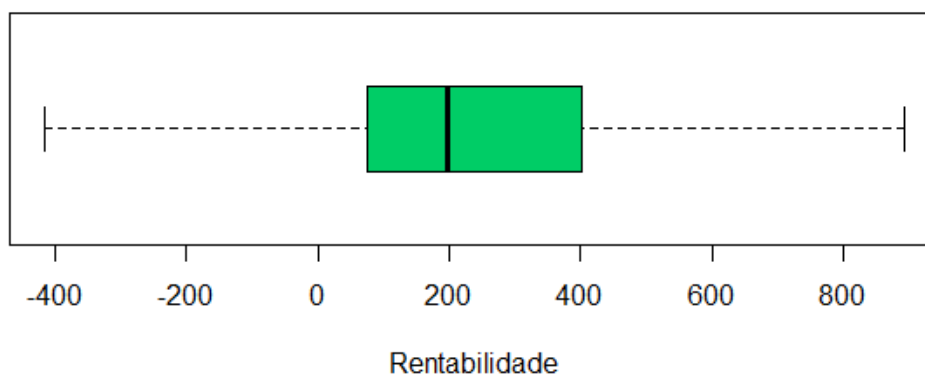


Figura 5.6 - Rentabilidade Gerada com barreira de outliers ligeiros

Para além dos valores máximos e mínimos já apresentados, é possível verificar que grande parte das observações se encontra entre 50 e 400, e mais concentrada acima do valor da mediana.

5.4.2. Rentabilidade Gerada - Resultados

Para avaliar a que nível um cliente é ou não rentável para a companhia de seguros será ajustada uma regressão Gama para ajustar os modelos e, os métodos Stepwise-backward e AIC para a escolha das variáveis que melhor explicam a variável resposta e que compõem o modelo final.

Com base em dados reais, que relacionam os prémios recebidos e os custos com os clientes, foi calculada a rentabilidade dos clientes que compõem a carteira da Seguros LOGO. Dada a existência de sinistros com custos bastante elevados, verificou-se a existência de valores de rentabilidade negativa. De forma a adequar os valores observados ao suporte de uma variável aleatória com distribuição Gama, efectuou-se uma transformação na variável dependente, caracterizada por $Y_i = R_i \times (-1) + 44.401$, de modo a que todas as observações da variável rentabilidade fossem positivas.

Na Figura 5.7 é possível observar o ajustamento dos dados, após a transformação atrás referida, a uma distribuição Gama.

Ajustamento dos Dados a uma Distribuição Gama

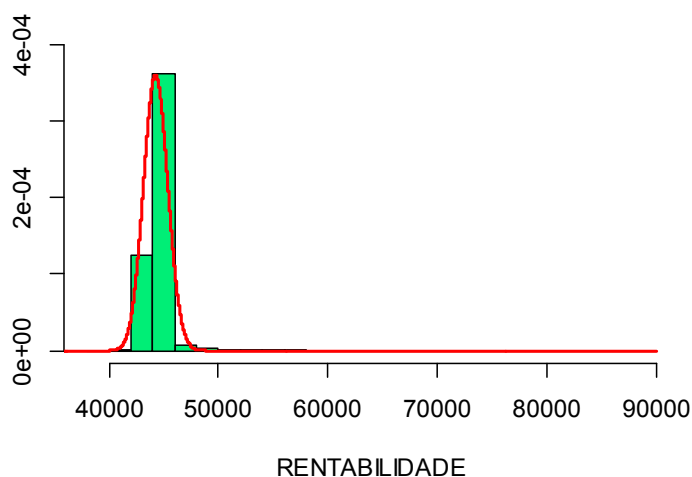


Figura 5.7 - Ajustamento dos dados a uma distribuição Gama

O modelo inicial criado para estimar quão rentável pode um cliente ser para a empresa, apresenta como variável resposta a rentabilidade do cliente, sendo esta, depois da transformação referida, uma variável contínua e positiva.

O conjunto de variáveis independentes inicialmente seleccionadas como tendo a capacidade de explicar a variável resposta é formado pelas mesmas covariáveis utilizadas na estimação da probabilidade de produção de lucro. Sendo, deste modo, o modelo completo composto pelas covariáveis: Dias de contrato, Canal, Produto, Fraccionamento, Bónus-Malus, Tempo de seguro, Idade do veículo, Tipo de veículo, Cilindrada, Combustível, Lugares, Prémio total, Sexo, Idade, Anos de carta, Zona, Quebra Isolada de Vidros, Veículo de Substituição, Franquia, Actos de Vandalismo e Fenómenos da Natureza, Indemnização Extra e Número de Sinistros nos últimos 5 anos.

O estudo que se segue foi efectuado com recurso ao *Software R*.

Modelo Nulo

De acordo com o que foi efectuado anteriormente iniciou-se a modelação ajustando o modelo nulo (em que todas as covariáveis são iguais a zero) obtendo-se os seguintes resultados:

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.13027	-0.00621	-0.00154	0.00120	0.78710

AIC: 3170955

Modelo Completo

Realizou-se a mesma análise para o modelo completo (contendo todas as covariáveis)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.12264	-0.00464	-0.00315	-0.00143	0.78628

AIC: 3162926

Os restantes outputs do R encontram-se na secção Anexos (9.2.4.1 - Rentabilidade Gerada - Resultados)

Comparação entre os dois modelos

Após a estimação dos dois modelos realizou-se a comparação dos mesmos com recurso ao *Teste de Razão de Verossimilhanças*, testando se o conjunto das covariáveis em análise é igual a zero ou estatisticamente significativo para explicar a variável dependente.

As hipóteses testadas são:

$$H_0: (\hat{\beta}_{DIAS_CONTRATOS}, \dots, \hat{\beta}_{SIN_5_ANOS}) = \mathbf{0}$$

vs

$$H_1: (\hat{\beta}_{DIAS_CONTRATOS}, \dots, \hat{\beta}_{SIN_5_ANOS}) \neq \mathbf{0}$$

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190126	99.468			
2	190059	95.288	67	4.1798	< 2.2e-16

Tendo por base o Teste de Wilks, verifica-se que o p-value obtido nos resultados acima é bastante pequeno (menor que 2×10^{-16}), e como tal rejeita-se a hipótese do conjunto das covariáveis ser igual a zero, havendo pelo menos uma covariável estatisticamente significativa. Resultado também verificado na análise da produção de lucros.

Para além dos resultados do teste de Wilks, os dois modelos estimados também podem ser comparados com base nos valores do Critério de Informação de Akaike (AIC). O modelo nulo apresenta um AIC de 3170955, enquanto o modelo completo apresenta um AIC menor, de 3162926, o que mais uma vez indica que o modelo completo, contendo todas as covariáveis, tem maior capacidade de explicativa que o modelo nulo.

Seleção do modelo através do Método Stepwise-Backward

O modelo completo apresentado no tópico anterior servirá com modelo inicial para a modelação através do método *Stepwise-Backward* (descrição da metodologia efectuada anteriormente no capítulo 4.5.1.2 - Backward Stepwise).

De acordo com este método, a primeira covariável a ser retirada do modelo é a que apresentar maior *p-value*, sendo neste caso a covariável Anos de Carta (ANOS_CARTA).

Após se retirar esta variável do modelo completo e se estimar os parâmetros do novo modelo, os dois modelos são comparados recorrendo-se ao *Teste de Razão de Verosimilhança*, a fim de se avaliar a hipótese do novo modelo ser estatisticamente mais significativo que o modelo sem a covariável, neste caso o modelo completo.

As hipóteses testadas são:

$$H_0: \hat{\beta}_{ANOS_CARTA} = \mathbf{0} \quad \text{vs} \quad H_1: \hat{\beta}_{ANOS_CARTA} \neq \mathbf{0},$$

obtendo-se:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190063	95.289			
2	190059	95.288	4	0.0010004	0.797

Analisando o resultado com base no *p-value* do teste conclui-se que, a covariável Anos de Carta deve ser retirada do modelo. Comparando o modelo completo, contendo todas as covariáveis, e o modelo sem esta covariável, tendo por base no *Teste de Wilks*, aceita-se a hipótese que defende que os parâmetros $\hat{\beta}_{ANOS_CARTA}$ são

iguais a zero ($p\text{-value} > 0,05$), não sendo estatisticamente significativos para a estimação da rentabilidade gerada. Passando, deste modo, o modelo sem a covariável Anos de Carta a ser o melhor modelo até ao momento.

Prossegue-se com a análise, considerando o modelo sem a variável e retirando-se a próxima covariável com maior $p\text{-value}$. Foram efectuados os mesmos procedimentos descritos anteriormente avaliando a exclusão ou não das restantes variáveis com menor significância estatística. Com base nos resultados obtidos, foram excluídas do modelo completo as covariáveis Anos de Carta, Indeminização Extra, Tipo de Veiculo, Canal e Sexo, permanecendo apenas o conjunto de covariáveis que melhor explicam a rentabilidade gerada.

Para uma análise mais detalhada, os resultados dos testes efectuados encontram-se nos Anexos, 9.2.4.3 - Selecção do modelo através do Método Stepwise-Backward.

O modelo obtido como sendo o melhor foi denominado Modelo Stepwise, sendo constituído pelas covariáveis: Dias de contrato, Produto, Fraccionamento, Bonus-malus, Tempo de Seguro, Idade do Veículo, Cilindrada, Combustível, Lugares, Prémio Total, Idade, Zona, Quebra Isolada de Vidros, Veículo de Substituição, Franquia, Actos de Vandalismo e Fenómenos da Natureza e Sinistros nos Últimos 5 anos.

Selecção das variáveis através do Critério de Informação de Akaike (AIC)

Como referido na estimação da Produção de lucro, outro método de selecção de modelos é através do Critério de Informação de Akaike (AIC), já descrito anteriormente.

O modelo obtido através da função acima mencionada, Modelo stepAIC, contém as mesmas covariáveis do Modelo Stepwise, mas com a diferença de não ter sido excluída a covariável Tipo de Veiculo.

Aplicando o Teste de Wilks, para comparar o Modelo StepAIC com o Modelo Completo obtém-se:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190064	95.273			
2	190057	95.271	7	0.0019577	0.8601

E quando se compara o mesmo modelo com o melhor modelo obtido através do Método Stepwise-Backward (Modelo Stepwise) os valores são:

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190067	95.290			

2 190070 95.295 -3 -0.0045213 0.05691

Ao analisar o p-value obtido verifica-se que o Modelo StepAIC é claramente melhor que o Modelo Completo, porém quando se compara o mesmo modelo com o Modelo Stepwise o valor do p-value é muito próximo de 0,05, passando o este a ser o melhor modelo, para um nível de confiança de 95%.

Como referido na modelação da Produção de Lucro, no Modelo Stepwise também existem casos de determinadas categorias, dentro das covariáveis, que não são significativamente diferentes da categoria definida como nível base do factor. De modo a simplificar o modelo, optou-se por agrupar as categorias não significativas à categoria que se encontra na base.

De modo a validar se as categorias com menor significância estatística devem ou não ser agrupadas ao nível base recorreu-se ao teste de Wald. Os resultados foram aferidos de acordo com o p-value obtido para os testes efectuados.

Após a realização dos testes necessários, as categorias agrupadas foram:

PRODUTO: LEASE+SAFE+MAX+TOPPING=TOPPING

FRACCIONAMENTO: ANUAL+SEMESTRAL=ANUAL

CILINDRADA: C1+C2+C5=C1

COMBUSTIVEL: D+E+M=D

LUGARES: 1+2+4+6+7+8+9=2

IDADE: I1+I3+I4+I8=I1; I5+I6+I7=5

ZONA: 1+2+4+5+6+9=1

Na Tabela 5.8 encontram-se os valores dos Critérios de Informação de Akaike dos principais modelos ajustados.

Modelo	AIC
Modelo Completo	3.162.926
Modelo Stepwise	3.162.917
Modelo StepAIC	3.162.914
Modelo Agrupado	3.162.917

Tabela 5.8 - Critério de Informação de Akaike (Rentabilidade)

Segundo os dados da tabela acima apresentada verifica-se que o modelo com melhor AIC é o Modelo StepAIC, não apresentando grandes diferenças face aos restantes modelos. O modelo Agrupado será, mais uma vez, o modelo utilizado na

restante análise da rentabilidade gerada, dado ser o modelo menos complexo, permitindo análises mais simples e resultados mais claros.

5.4.3. Rentabilidade Gerada - Análise dos Resíduos

De acordo com o efectuado na análise da Produção de Lucro, para a Rentabilidade Gerada também se procedeu à representação gráfica dos resíduos.

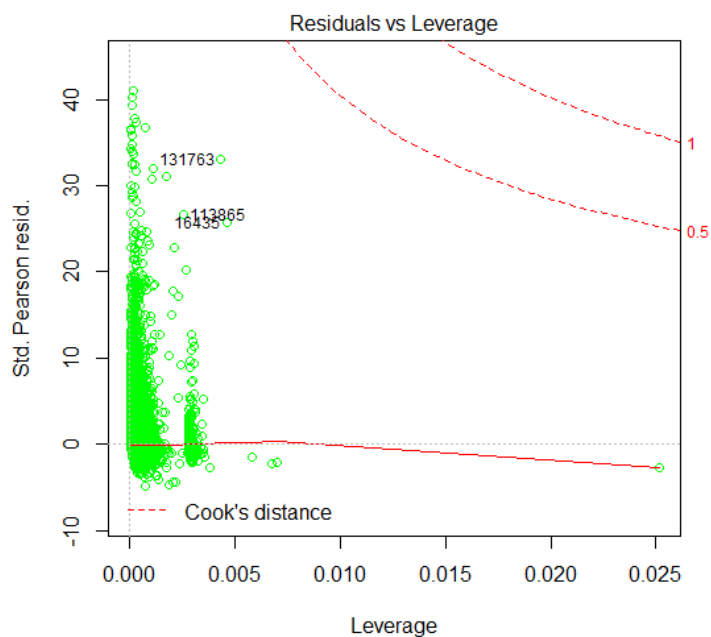


Figura 5.8 - Resíduos de Pearson Padronizados vs Alavancagem

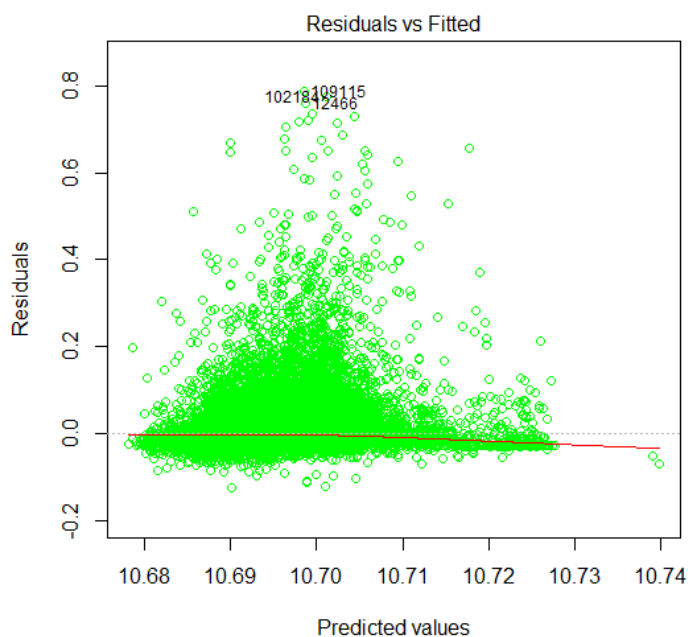


Figura 5.9 - Desvios Residuais vs Valores Ajustados

Observando a Figura 5.9 verifica-se que os pontos da sua grande maioria apresentam valores baixos de alavancagem. Contudo, existe pelo menos um ponto com um valor superior que poderia representar uma forte influência, aproximando de si a linha da regressão, diminuindo o valor do resíduo associado à sua observação. Mas o mesmo não se verifica, pois o ponto em causa apresenta uma reduzida Distância de Cook.

Com base na Figura 5.9 verifica-se que grande parte das observações pode ser descrita através do modelo ajustado, no entanto, existem ainda algumas observações cujo desvio residual indica que o modelo ajustado não se encontra totalmente adequado à explicação das mesmas.

5.5. ESTIMAÇÃO DA RENTABILIDADE GERADA

Após o estudo estatístico dos dados e aplicação de métodos para selecção das covariáveis foi apurado o modelo final que melhor explica a variável resposta (Rentabilidade Gerada). Mais uma vez o Modelo Agrupado foi o modelo escolhido como modelo final, dado ser o modelo mais simplificado.

O modelo final é composto pelas seguintes covariáveis independentes Dias de Contrato, Produto, Fraccionamento, Bónus Malus, Tempo de Seguro, Idade do Veículo, Cilindrada, Combustível, Lugares, Prémio Total, Idade, Zona, Quebra Isolada de Vidros, Veículo de Substituição, Franquia, Actos de Vandalismo e Fenómenos da natureza e Sinistros nos Últimos 5 anos.

Do ajustamento do modelo final resultam os coeficientes β_i que serão utilizados para a estimação rentabilidade gerada para a empresa.

As estimativas dos coeficientes e o modelo utilizados podem ser consultados na secção Anexos (9.1.1.8 - Estimação da Rentabilidade Gerada).

Na expressão que define o modelo estão expressas todas as categorias que compõem cada variável, mas é de realçar que uma categoria de cada variável categórica está associada ao parâmetro β_0 .

O conjunto de tais características será definido como o perfil do *Cliente Padrão*, sendo, mais concretamente, definido por:

PRODUTO	LIGHT
FRACCIONAMENTO	ANUAL
TEMPO_SEGURO	TS3
IDADE_VEICULO	IV1
CILINDRADA	C1
COMBUSTIVEL	D
LUGARES	L2
IDADE	I1
ZONA	Z1
QIV	N
VS	N
FRANQUIA	F0
AV_FN	N
SIN_5_ANOS	SIN1

Tabela 5.9 - Características do *Cliente Padrão* (Rentabilidade)

De acordo com o procedimento efectuado na Produção de Lucro e de modo a simplificar a análise do modelo estimado, na Tabela 5.10 encontram-se representados alguns exemplos da estimação da rentabilidade que clientes com determinadas características podem gerar para a empresa.

	Cliente A	Cliente B	Cliente C	Cliente D	Cliente E
DIAS_CONTRATO	500	800	1000	750	750
PRODUTO	TOPPING	MOTO	TOPPING	LIGHT	LIGHT
FRACCIONAMENTO	ANUAL	MENSAL	ANUAL	TRIMESTRAL	TRIMESTRAL
BONUS_MALUS	50	80	50	60	60
TEMPO_SEGURO	TS3	TS4	TS2	TS5	TS5
IDADE_VEICULO	IV4	IV4	IV2	IV3	IV3
CILINDRADA	C4	C1	C3	C3	C3
COMBUSTIVEL	D	G	G	D	D
LUGARES	L5	L2	L5	L5	L5
PREMIO_TOTAL	250	120	500	150	150
IDADE	I3	I2	I2	I5	I5
ZONA	Z7	Z8	Z1	Z3	Z3
QIV	S	N	N	N	N
VS	S	N	S	S	S
FRANQUIA	2	0	4	0	0
AV_FN	N	N	N	N	N
SIN_5_ANOS	SIN1	SIN2	SIN3	SIN1	SIN3
$(e_i^Y - 44.400,11) * (-1)$	481€	215€	46€	262€	-256 €
IC (95%)	[476€;487€]	[210€;221€]	[40€;51€]	[267€;257€]	[-261€;-250€]

Tabela 5.10 - Rentabilidade Gerada – exemplo

Na tabela acima, pode observar-se o resultado da estimação da rentabilidade gerada por clientes exemplo.

De acordo com dados apresentados espera-se que um cliente que apresente as mesmas características que o Cliente A gere uma rentabilidade de 481€ para a empresa, podendo afirma-se com 95% de confiança que a rentabilidade gerada se encontra entre os 476€ e os 487€. Estes valores indicam que um individuo que apresente estas características provavelmente será um cliente rentável para a seguradora.

Por sua vez, um cliente com características idênticas às do cliente C espera-se que retorne uma rentabilidade positiva, no entanto, com um valor baixo. Uma carteira

com um elevado número de clientes com tais características poderá não ser benéfico para a empresa uma vez que os custos com sinistros não representam a totalidade dos custos que a seguradora tem a seu cargo. Deste modo, os ganhos que terá com estes clientes poderão não ser suficientes para se manter rentável e no limite solvente.

De acordo com o apresentado para a produção de lucro, os clientes D e E apresentam as mesmas características com excepção da variável Número de Sinistros nos últimos 5 anos. A alteração desta variável, da categoria SIN1 para SIN3, faz com que a rentabilidade esperada de um cliente passe de positiva (262€) para negativa (-256€).

A variável Número de Sinistros nos últimos 5 anos é claramente significativa na estimação da rentabilidade gerada pelos clientes, porém não pode ser interpretada de modo isolado. O Cliente E e o Cliente C apresentam o mesmo o mesmo valor para a variável em causa, mas o Cliente E, mesmo que baixa, apresenta uma rentabilidade positiva.

Para a estimação da rentabilidade gerada pelos clientes, não basta analisar uma ou outra variável isoladamente, é necessário ter em conta o conjunto das várias características do cliente, do objecto seguro e do nível de risco a cobrir.

6. CONCLUSÕES

Uma seguradora automóvel no decorrer da sua actividade, para além de cobrir riscos de sinistralidade deve ser uma empresa rentável e que gere lucros. Porém com a subida da taxa de sinistralidade as companhias têm visto a sua rentabilidade diminuir, pois a percentagem dos prémios cobrados que está a ser utilizada para cobrir os custos com sinistros é cada vez maior. Tal facto reflecte a necessidade cada vez maior de conhecer a rentabilidade dos clientes para o cálculo de uma tarifa bem ajustada ao risco, pois os sinistros não correspondem à totalidade dos riscos a que a seguradora se encontra exposta, nem à totalidade dos custos da mesma.

Cada vez mais, há uma grande necessidade por parte das companhias de seguros de investir mais recursos numa boa gestão de risco e não se focar apenas na dimensão da carteira. Mais importante que angariar clientes é angariar e reter clientes rentáveis.

Neste projecto, com o auxílio dos MLG, foram identificadas as variáveis que maior impacto têm na produção ou não de lucros e geração de rentabilidade dos clientes. Através dos resultados obtidos é possível conhecer as características de um cliente rentável e menos rentável. O conhecimento de tais características tem uma importância relevante para um bom desempenho da actividade seguradora, pois permite ajustar a tarifa de acordo com os níveis reais de risco dos vários perfis de clientes e conseqüentemente a constituição e manutenção dos requisitos de capital necessários, permitindo assumir todas as suas responsabilidades sem por em causa a solvência da companhia.

O conhecimento das características dos clientes potencialmente rentáveis pode também ser utilizado como vantagem competitiva face à concorrência. Permite à empresa a elaboração de campanha de angariação e retenção para um determinado público-alvo, possibilitando a constituição de uma carteira mais rentável (com níveis de risco mais baixos).

7. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Ao elaborar este trabalho a principal limitação encontrada foi relativamente ao tamanho da base de dados analisada. Quando os dados foram retirados a empresa em estudo apenas se encontrava em actividade há três anos, não reunindo um número muito elevado de registos. Este facto pode resultar num enviesamento maior dos resultados apresentados, pois os métodos utilizados são normalmente aplicados a universos de maior dimensão.

Após uma gestão cuidada da carteira de clientes e de uma tarifa bem calculada e ajustada ao real risco a que a empresa se encontra exposta de modo constituir uma rentabilidade estável, uma possível análise futura poderia estar relacionada com a procura. Esta análise permitiria à empresa conhecer o perfil de clientes que efectua simulações mas que não as converte em contrato, optando possivelmente por um produto da concorrência, uma vez que o seguro automóvel é obrigatório por lei.

“Será que a LOGO está a perder possíveis clientes rentáveis ou apenas os que apresentam níveis de risco elevado?”

8. BIBLIOGRAFIA

- Anónimo. (2011). Apontamentos Estatística Aplicada. *Diagramas Caixa e Bigódes*, pp. 30-31. Retirado de http://pt.scribd.com/doc/53330501/26/Diagramas-de-Caixa-e-Bigodes#outer_page_26 [acedido em Outubro de 2011]
- Anónimo. (2010). *Tradicionais ou "low cost"?* Auto Hoje, (Desporto e Veículos), p.14. Retirado da Intranet a 15 de Abril de 2010
- Autoridade Nacional Segurança Rodoviária. (2012). *Dados Sinistralidade*. Retirado de <http://www.ansr.pt/Default.aspx?tabid=344> [acedido em Agosto de 2013]
- Associação Portuguesa de Seguradores. (2010). *Dados do seguro automóvel de final de 2010*. Retirado de <http://www.apseguradores.pt/Site/Content.jsf?ContentId=1831> [acedido Setembro de 2011]
- Associação Portuguesa de Seguradores. (2010). *Estatística Automóvel*. Retirado de <http://www.apseguradores.pt/DocumentCatalog/DocumentLink2.jsf?DocumentId=6484> [acedido em Setembro de 2011]
- Associação Portuguesa de Seguradores. (2010). *O seguro automóvel em Portugal*. Retirado de http://www.vidaeconomica.pt/users/0/39/Seguros_2cb2a8946435eff4c39a07f283a0f2ac.pdf [Acedido em Julho de 2011]
- Borginho, Hugo. (2001). *Metodologias Estocásticas de Estimação de Provisões Para Sinistros* (Estágio Curricular). Lisboa.
- Borginho, Hugo. (2010). *Basileia II*. (Apontamentos de aulas). Lisboa
- Carvalho, Raquel. (2011). Seguros. *Diário Económico (Suplemento)*, 5137.
- Crawley, Michel J. (2007). *The R Book*. Retirado de http://www.kharms.biology.lsu.edu/CrawleyMJ_TheRBook.pdf. [acedido em Agosto 2012].
- Gomes, Ana Santos (2010). *Preço do seguro automóvel sobe em 2011*, pp, 4-5. Retirado de http://www.vidaeconomica.pt/users/0/39/Seguros_2cb2a8946435eff4c39a07f283a0f2ac.pdf [acedido em Setembro 2012].
- Guerreiro, Tiago Miguel. (2008). *Análise da sinistralidade rodoviária em Portugal*. Retirado de <https://dspace.ist.utl.pt/bitstream/2295/236519/1/Dissertacao%20com%20anexos.pdf> [acedido em Setembro 2011].
- Instituto de Seguros de Portugal. (2007). Análise da produtividade do sector segurador. *Relatório do Sector Segurador e Fundos de Pensões*, 1645-4049, pp. 237-350
- Instituto de Seguros de Portugal. (2012). *Parque automóvel seguro*. Retirado de <http://www.isp.pt/NR/exeres/7D383D46-9431-416E-98C7-395B0A9E7080.htm> [acedido em Agosto de 2014]

- Instituto de Seguros de Portugal. (2012). Prémios. Retirado de <http://www.isp.pt/NR/exeres/A6E856FF-AD45-40DD-83C7-8614A910D835.htm> [acedido em Agosto de 2014]
- Leitão, C.R.S. (2002) *Investigação da Rentabilidade de Clientes: um estudo no sector hoteleiro do Nordeste* (Tese de Mestrado). Retirado de [http://www.biblioteca.sebrae.com.br/bds/BDS.nsf/EB18DA9A67DE640083256F800049228B/\\$File/NT000A2EAE.pdf](http://www.biblioteca.sebrae.com.br/bds/BDS.nsf/EB18DA9A67DE640083256F800049228B/$File/NT000A2EAE.pdf) [acedido em Abril de 2010]
- Leitão, C.R.S. (2007) *Rentabilidade de Clientes: Um estudo sob a ótica do ABM (Activity Based Management)*. Retirado de [http://www.biblioteca.sebrae.com.br/bds/BDS.nsf/1ED74DAAE185328883256F7100642D26/\\$File/NT000A2482.pdf](http://www.biblioteca.sebrae.com.br/bds/BDS.nsf/1ED74DAAE185328883256F7100642D26/$File/NT000A2482.pdf) [acedido em Abril de 2010]
- Marketeer. (2010). Segmento Low Cost, LOGO... logo resultou! *Marketeer*, pp.140-143. Retirado da Intranet a Abril de 2010
- Pontes, Marco. (2010). *Solvencia II – Uma abordagem didática*. Retirado de <http://marcoponteslgpconsulting.blogspot.pt/2010/09/solvencia-ii-uma-abordagem-didatica.html> [acedido em Setembro 2011]
- Portugal. (2007). Decreto-lei nº.291/2007. Diário da República nº160, série 1 de 2007-08-21 [Obrigação de Seguro], pp.5488-5489
- Prudente, Adrea Adrade. (2009). *Modelos Não-Linearres de regressão: Alguns aspectos de teoria assintótica*. Retirado de http://www.pgbiom.ufrpe.br/dissertacoes/2009/dissertacao_final_andrea_andrade_prudente.pdf [acedido em Junho de 2010]
- Real, Pedro Corte. (2009/2010). *Solvência II* (apontamentos de aulas). Lisboa
- Real, Pedro Corte. (2009/2010). *Gestão de Risco e Contro Interno. Capital Económico* (apontamentos de aulas). Lisboa
- Santos, Susete Tomás dos. (2008). *Construção de uma tarifa de responsabilidade civil automóvel*. Retirado de http://run.unl.pt/bitstream/10362/2447/1/Santos_2008.pdf. [acedido em Agosto de 2012]
- Seguros Logo SA. (2008). *Logo revoluciona o mercado segurador*. Lisboa: Seguros Logo SA. Retirado da Intranet a Agosto de 2009
- Seguros Logo SA. (2008). *Tudo sobre a Logo*. Lisboa: Seguros Logo SA. Retirado da Intranet a Agosto de 2009
- Seguros Logo SA. (2009). *Formação*. Lisboa: Seguros Logo SA. Retirado da Intranet a 03 de Agosto de 2009
- Seguros Logo SA. (2009). *Manual_oferta_kiosks_05_11_09*. Lisboa: Seguros Logo SA. Retirado da Intranet a Janeiro de 2010
- Turkman, M. Antónia Amaral e Silva, Giovanni Loiola. *Modelos Lineares Generalizados – Da teoria à prática*. Retirado de <http://docentes.deio.fc.ul.pt/maturkman/mlg.pdf> [acedido em Agosto de 2012]
- Vale, Catarina Alexandra Lopes do. *Modelação e Estimação do Risco de Crédito – Estudo de uma Carteira*. Retirado de <http://ferrari.dmat.fct.unl.pt/personal/mle/PUBL-rdf/Teses/MSc-CatarinaVale.pdf> [acedido em Agosto de 2014].

9. ANEXOS

9.1.1.1. *Descrição de Coberturas*

- **Assistência em viagem:** em caso de acidente ou avaria, a seguradora garante o reboque do veículo, o transporte de pessoas e bens e o fornecimento de outro veículo até ao final da viagem.
- **Protecção jurídica:** cobre os custos de um advogado que represente os interesses do segurado e as despesas dos processos judiciais ou administrativos.

Ainda no âmbito dos seguros de responsabilidade civil, é possível subscrever coberturas adicionais como:

- **Quebra Isolada de Vidros:** garante a reparação em caso de quebra isolada dos seguintes vidros: pára-brisas, óculo traseiro e vidros laterais, mas que não tenha sido causada por danos cobertos por outra cobertura, como é o caso de quebras resultantes de actos de vandalismo ou fenómenos naturais (tempestades, quedas de árvores, etc.)
- **Protecção de ocupantes:** ao subscrever esta cobertura, a seguradora responsabiliza-se pelas despesas de tratamento do segurado bem como dos restantes ocupantes da viatura até determinado limite de capital.
- **Privação temporária de uso:** garante o pagamento de uma indemnização pelos prejuízos resultantes da impossibilidade de utilização do veículo seguro durante um determinado período de tempo.
- **Veículo de substituição:** em caso de acidente ou avaria do objecto seguro, a seguradora garante uma viatura de acordo com determinadas condições de utilização.

Na componente de Danos próprios as coberturas possíveis de subscrever são:

- **Choque, colisão ou capotamento:** A companhia de seguros garante o pagamento dos danos (até ao limite do capital seguro), no caso de choque, colisão ou capotamento da viatura, mesmo que o condutor seja responsável pelo sinistro.
- **Incêndio, raio ou explosão:** garante o pagamento dos prejuízos causados no veículo seguro, que tenham sido causados por incêndio, explosão casual ou queda de raios, independentemente do veículo se encontrar parado ou em movimento.

- **Furto ou roubo:** Cobre o pagamento de danos causados no veículo, ou a sua perda, derivados de furto ou roubo ou tentativa de tais actos.
- **Actos de vandalismo:** Garante o pagamento ou reparação de danos provocados por acção humana, directa e voluntária no veículo seguro.
- **Fenómenos da natureza:** ao subscrever esta cobertura o segurado garante que a seguradora se encarrega do pagamento dos danos causados por fenómenos naturais como ciclones, terramotos, trombas de água e enxurradas.

9.1.1.2. Tipos de Risco

- **Risco de mercado** – Impacto das variáveis financeiras como a variação das cotações de acções, da taxa de juro, da taxa de cambio, do valor de imóveis, do *credit spread*, que influenciam o cálculo da melhor estimativa.
- **Risco de contraparte** – Incerteza da contraparte de um determinado contrato cumprir com as suas obrigações financeiras (resseguro).
- **Risco operacional** – Perdas relacionadas com falhas ao nível de sistemas internos, de recursos humanos, risco legal, acontecimentos externos. Todos estes riscos são bastante difíceis de quantificar devido à falta ou reduzido número de dados históricos.
- **Risco de Reservas** – Risco associado ao montante necessário para pagamento de sinistros.
- **Risco de Prémios** – Risco de que o volume de prémios não seja suficiente para cumprimento de todas as responsabilidades da seguradora.

9.1.1.3. Oferta Logo

- **Logo Light** – Seguro de Responsabilidade Civil Automóvel direccionado para os clientes que apenas procuram cumprir o estabelecido pela obrigatoriedade da lei, optando pelo produto mais simples, sendo o mais barato, contendo somente as coberturas obrigatórias (Responsabilidade Civil, Assistência em Viagem e Assistência Jurídica)
- **Logo Topping** – Seguro de Responsabilidade Civil Automóvel, mas que contém uma cobertura que cobre o risco dos ocupantes e a possibilidade de subscrever a cobertura de Quebra Isolada de Vidros.
- **Logo Safe** – Seguro de Responsabilidade Civil Automóvel, mas que para além das coberturas apresentadas no *Logo Topping*, permite ao cliente

cobrir o risco de Furto ou Roubo. Este produto é direccionado para clientes que se preocupam com a segurança da viatura.

- **Logo Max** – Seguro de Responsabilidade Civil e Danos Próprios, direccionado para clientes que pretendem cobrir todo o tipo de risco, tanto os que afectam terceiros como a sua viatura. É um produto normalmente escolhido por clientes que possuem carros relativamente novos.
- **Logo Moto** – Seguro de Responsabilidade Civil Automóvel destinado a motociclos ou ciclomotores composto pelas coberturas de Responsabilidade Civil, Assistência em Viagem e Assistência Jurídica.

9.1.1.4. Variáveis Utilizadas

VARIÁVEL	CATEGORIA	DESCRIÇÃO
DIAS_CONTRATO	(VARIÁVEL DISCRETA)	Tempo de Permanência na Empresa
CANAL	CALL_CENTER INTERNET PARCEIROS	Contratos por Telefone Contratos por Internet Contratos através de entidades Externas
PRODUTO	LIGHT TOPPIN SAFE MAX MOTO LEASE	Produto de Responsabilidade Civil Produto de Responsabilidade Civil Produto de Responsabilidade Civil Produto de Danos Próprios Produto Exclusivo para motociclos e ciclomotores Produto de Danos Próprios
FRACCIONAMENTO	MENSAL TRIMESTRAL SEMESTRAL ANUAL	Pagamento em 12 prestações Pagamento em 4 prestações Pagamento em 2 prestações Pagamento completo
BONUS_MALUS	(VARIÁVEL DISCRETA)	Nível de Bónus ou agravamento
TEMPO_SEGURO	TS1; TS2; TS3; TS4; TS5	Tempo de Seguro Automóvel
IDADE_VEICULO	IV1; IV2; IV3; IV4	Idade do Objecto Seguro
TIPO_VEICULO	B C L M	Ligeiro ou Comercial Comercial Ligeiro Moto
CILINDRADA	C1 a C5	Cilindrada do Objecto Seguro
COMBUSTIVEL	D E G M	Veículo a Diesel Veículo Eléctrico Veículo a Gasolina Veículo Híbrido
LUGARES	L1 a L9	Número de Lugares do veículo
PREMIO_TOTAL	(VARIÁVEL DISCRETA)	Prémio cobrado ao cliente pelo seguro subscrito
SEXO	F M	Segurado do sexo feminino Segurado do sexo masculino
IDADE	I1 a I8	Idade do Segurado
ANOS_CARTA	TC1 a TC5	Tempo de Carta do segurado
ZONA	Z1 a Z9	Zona de Residência do segurado
QIV	N;S	Cobertura Quebra Isolada de Vidros
VS	N;S	Cobertura Veículo de Substituição
FRANQUIA	F0, F2, F4	Percentagem da franquia estabelecida no momento da subscrição
AV_FN	N;S	Cobertura Actos de Vandalismo e Fenómenos da Natureza
IE	N;S	Cobertura Indemnização Extra
SIN_5_ANOS	SIN1 a SIN4	Número de Sinistros nos últimos 5 anos antes do actual contrato

R	1;0	Produção de Lucros
RENT_POSIT	(VARIABLE CONTINUA)	Rentabilidade do cliente

Tabela 9.1 - Variáveis e respectivas categorias

Produção de Lucros (R): Variável *dummy* que assume o valor 1 no caso de o cliente ter originado lucro para a empresa e o valor 0 em caso contrário.

Rentabilidade Gerada (RENT_POSIT): Variável categórica onde está expresso o montante corresponde à diferença entre o valor pago pelo cliente relativo ao prémio de seguro e os custos da seguradora em caso de sinistro.

Dias de Contrato (DIAS_CONTRATO): Variável numérica que indica há quantos dias o contrato se encontra activo. Os contratos têm a duração de um ano, mesmo que o cliente renove o contrato este é novamente tarifado podendo o cliente apresentar outras características como é o caso da idade, tempo de carta, entre outras.

Canal (CANAL): Variável categórica que classifica o canal de distribuição utilizado para a conversão do contrato. O cliente pode ter subscrito o contrato pela Internet, Call Center ou através de Parceiros da Seguros Logo.

Produto (PRODUTO): Variável categórica que identifica o produto que o cliente subscreveu ao fazer o contrato com a companhia (descrição dos produtos no tópico *Oferta Logo*). Os produtos abrangidos pela análise deste projecto são o Logo Light (LIGHT), Logo Topping (TOPPING), Logo Safe (SAFE), Logo Max (MAX), Logo Moto (MOTO) e Logo Lease (LEASE).

Fraccionamento (FRACCIONAMENTO): Variável categórica que indica o fraccionamento escolhido pelo cliente para pagamento do prémio de seguro. o pagamento pode ser feito de uma só vez sendo o fraccionamento anual (ANUAL) ou repartido em duas parcelas (SEMESTRAL), quatro parcelas (TRIMESTRAL) ou ainda em 12 parcelas (MENSAL).

Bonus-Malus (BONUS_MALUS): Variável numérica que indica o nível de risco do cliente atribuindo-lhe um nível que pode corresponder a um desconto ou agravamento de acordo com o seu histórico de sinistralidade. O nível de Bonus-Malus é actualizado sempre que existe retarificação.

Tempo de Seguro (TEMPO_SEGURO): Variável categórica que indica há quanto tempo o segurado possui seguro automóvel em seu nome, independentemente da companhia de seguros. Esta variável é resultado do agrupamento dos valores possíveis em 5 classes onde a classe 1 (TS1) corresponde a segurados com menos anos de seguro e a classe 5 (TS5) a segurados com maior experiência de seguro, (por confidencialidade de dados não serão apresentados os intervalos reais utilizados).

Idade do Veículo (IDADE_VEICULO): Variável categórica que indica a idade do veículo à data do início do contrato. Esta variável é resultado do agrupamento dos valores possíveis em 4 classes onde a classe 1 (IV1) corresponde veículos mais recentes e a classe 4 (IV4) a veículos mais antigos, (por confidencialidade de dados não serão apresentados os intervalos reais utilizados).

Tipo de Veículo (TIPO_VEICULO): Variável categórica que classifica o objecto seguro de acordo com a sua classe e tipo (classificação de acordo com a Eurotax). A seguradora em estudo apenas comercializa automóveis ligeiros de passageiros (L), ligeiros de mercadorias de uso pessoal (C), que segundo Monteiro (2006) são veículos que no máximo podem ter 3500kg e não podem transportar mais de nove passageiros incluindo o condutor, motociclos ou ciclomotores (M). Existe ainda o tipo ligeiro de passageiros ou mercadorias (B).

Cilindrada do veículo (CILINDRADA): Variável categórica que indica a cilindrada do objecto seguro. Esta variável é resultado do agrupamento dos valores possíveis em 5 classes onde a classe 1 (C1) corresponde a veículos com menor cilindrada e a classe 5 (C5) a veículos com maior cilindrada, (por confidencialidade de dados não serão apresentados os intervalos reais utilizados).

Combustível (COMBUSTIVEL): Variável categórica que indica o tipo de combustível que o veículo seguro utiliza. Os veículos seguros pela Logo podem ser a gasolina (G), diesel (D), híbridos (M) ou eléctricos (E).

Lugares (LUGARES): variável categórica que define o número de lugares do objecto seguro. Pode assumir valores entre 2 e 9 lugares.

Prémio Total (PREMIO_TOTAL): Variável numérica onde está expresso o montante (€) que o tomador de seguro pagou à companhia para segurar o seu veículo ao abrigo do contrato subscrito com a seguradora.

Sexo (SEXO): Variável categórica que identifica o género do condutor habitual, masculino (M) ou feminino (F).

Idade (IDADE): Variável categórica que indica a idade do condutor habitual à data do início do contrato. Esta variável é resultado do agrupamento dos valores possíveis em 8 classes onde a classe 1 (I1) corresponde a clientes mais jovens e a classe 8 (I8) a clientes mais velhos, (por confidencialidade de dados não serão apresentados os intervalos reais utilizados).

Tempo de Carta (ANOS_CARTA): Variável categórica que indica há quanto tempo o segurado possui carta de condução. Esta variável é resultado do agrupamento dos valores possíveis em 5 classes onde a classe 1 (TC1) corresponde a segurados com

menos anos de carta 5 (TC5) a segurados com maior experiencia, (por confidencialidade de dados não serão apresentados os intervalos reais utilizados).

Zona (ZONA): Variável categórica que classifica a zona de residência do condutor habitual de acordo com o seu código-postal (assume valores entre 1 e 9 de acordo com critérios estabelecidos pela seguradora).

Quebra Isolada de Vidros (QIV): Variável categórica que assume a designação ‘S’ caso o contrato subscrito incluia a cobertura Quebra Isolada de Vidros ou ‘N’ em caso contrário.

Veículo de Substituição (VS): Variável categórica que assume a designação ‘S’ caso o contrato subscrito incluia a cobertura Veículo de Substituição ou ‘N’ em caso contrário.

Franquia (FRANQUIA): Variável categórica que identifica a percentagem do montante referente ao custo de reparação que fica a cargo do tomador do seguro em caso de sinistro. A percentagem mínima é zero caso o tomador opte por deixar a totalidade do custo do sinistro a cargo da seguradora. No caso de optar por deixar parte do custo a seu cargo pode optar por 2 ou 4%.

Actos de Vandalismo e Fenómenos da Natureza (AV_FN): Variável categórica que assume a designação ‘S’ caso o contrato subscrito incluia a cobertura Actos de Vandalismo e Fenómenos da Natureza ou ‘N’ em caso contrário.

Indemnização Extra (IE): Variável categórica que assume a designação ‘S’ caso o contrato subscrito incluia a cobertura Indemnização Extra ou ‘N’ em caso contrário.

Coberturas Facultativas (COB_FACULT): Variável categórica que indica o número de coberturas opcionais que o tomador incluiu no contrato subscrito. Pode assumir valores entre 0 e 3.

Sinistros nos últimos 5 anos (SIN_5_ANOS): Variável categórica que indica o número de sinistros com culpa que o clientes teve antes de subscrever o actual seguro. Esta variável é resultado do agrupamento dos valores possíveis em 4 classes onde a classe 1 (SIN1) corresponde a segurados com menos sinistros e a 4 (SIN4) a segurados com mais sinistros, (por confidencialidade de dados não serão apresentados os intervalos reais utilizados).

9.1.1.5. *Análise estatística das variáveis Produção de Lucro*

	CANAL				FRACCIONAMENTO		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
Call_Center	30%	31%	29%	MENSAL	4%	4%	4%

Internet	32%	28%	33%	TRIMESTRAL	32%	36%	32%
Parceiros	38%	40%	38%	SEMESTRAL	24%	26%	24%
				ANUAL	39%	34%	40%
	PRODUTO				CILINDRADA		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
LIGHT	15%	16%	15%	C1	21%	15%	21%
TOPPING	56%	61%	56%	C2	47%	51%	47%
SAFE	7%	7%	7%	C3	27%	28%	26%
MAX	9%	12%	9%	C4	4%	4%	4%
MOTO	12%	4%	13%	C5	1%	1%	1%
LEASE	0%	0%	0%				
	TEMPO_SEGURO				COMBUSTIVEL		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
TS1	2%	1%	2%	D	41%	35%	42%
TS2	11%	11%	11%	E	0%	0%	0%
TS3	6%	6%	6%	G	59%	65%	58%
TS4	6%	7%	6%	M	0%	0%	0%
TS5	75%	76%	75%				
	IDADE_VEICULO				TIPO_VEICULO		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
IV1	2%	2%	2%	B	2%	2%	2%
IV2	20%	20%	20%	C	7%	7%	7%
IV3	65%	67%	65%	L	79%	87%	78%
IV4	13%	12%	13%	M	13%	4%	13%
	LUGARES				ZONA		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
L1	2%	1%	1%	1	7%	8%	7%
L2	17%	10%	18%	2	17%	20%	17%
L3	1%	1%	1%	3	22%	23%	22%

L4	4%	4%	4%	4	14%	16%	14%
L5	73%	81%	72%	5	19%	17%	19%
L6	0%	0%	0%	6	11%	10%	11%
L7	2%	2%	2%	7	6%	4%	6%
L8	0%	0%	0%	8	2%	1%	2%
L9	0%	0%	0%	9	2%	1%	2%
	IDADE				ANOS_CARTA		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
I1	0%	0%	0%	AC1	0%	0%	0%
I2	1%	1%	1%	AC2	2%	3%	2%
I3	8%	8%	7%	AC3	4%	6%	4%
I4	16%	15%	16%	AC4	6%	6%	6%
I5	39%	37%	39%	AC5	88%	85%	88%
I6	21%	22%	21%				
I7	11%	11%	10%				
I8	5%	5%	5%				
	SEXO				FRANQUIA		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
F	25%	29%	25%	0	95%	96%	95%
M	75%	71%	75%	2	3%	2%	3%
				4	2%	2%	2%
	QIV				VS		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
N	76%	75%	76%	N	80%	76%	81%
S	24%	25%	24%	S	20%	24%	19%
	AV_FN				IE		
	TOTAL	N_PROD_L	PROD_L		TOTAL	N_PROD_L	PROD_L
N	93%	90%	93%	N	91%	89%	91%
S	7%	10%	7%	S	9%	11%	9%

	SIN_5_ANOS		
	TOTAL	N_PROD_L	PROD_L
SIN1	94%	86%	94%
SIN2	6%	11%	5%
SIN3	1%	2%	1%
SIN4	0%	1%	0%

Tabela 9.2 – Produção de Lucros - Análise Preliminar

9.1.1.6. Análise estatística das variáveis Rentabilidade Gerada (5.1.2)

CANAL		FRACCIONAMENT		TIPO_VEICULO	
CALL_CENTER	165,02 €	MENSAL	112,21 €	B	163,50 €
INTERNET	132,26 €	TRIMESTRAL	34,34 €	C	93,40 €
PARCEIROS	96,92 €	SEMESTRAL	130,37 €	L	135,21 €
		ANUAL	186,42 €	M	99,42 €
PRODUTO		CILINDRADA		ANOS CARTA	
LIGHT	82,31 €	C1	110,91 €	TC1	-68,96 €
TOPPING	122,75 €	C2	132,41 €	TC2	57,79 €
SAFE	151,63 €	C3	141,41 €	TC3	95,40 €
MAX	266,38 €	C4	107,19 €	TC4	103,16 €
MOTO	97,78 €	C5	68,92 €	TC5	133,57 €
LEASE	346,82 €				
TEMPO_SEGURO		COMBUSTIVEL		IDADE_VEICULO	
TS1	17,98 €	D	137,20 €	IV1	226,34 €
TS2	123,31 €	E	305,68 €	IV2	217,59 €
TS3	157,11 €	G	122,54 €	IV3	107,70 €
TS4	120,32 €	M	-251,23 €	IV4	77,88 €
LUGARES		ZONA		IDADE	
L1	98,51 €	1	94,62 €	I1	3,63 €

L2	107,84 €	2	115,75 €	12	112,45 €
L3	21,54 €	3	132,31 €	13	112,84 €
L4	181,31 €	4	122,22 €	14	111,72 €
L5	130,54 €	5	126,81 €	15	133,09 €
L6	123,12 €	6	143,19 €	16	124,53 €
L7	212,27 €	7	163,59 €	17	134,42 €
L8	202,28 €	8	178,58 €	18	188,49 €
L9	76,28 €	9	140,72 €		
SEXO			FRANQUIA		
F	152,69 €	F0	116,52 €		
M	120,29 €	F2	439,54 €		
		F4	286,61 €		
	QIV	VS	AV_FN	IE	
N	120,48 €	119,20 €	116,45 €	127,52€	
S	153,88 €	165,89 €	282,98 €	138,07€	
SIN_5_ANOS					
SIN1	144,58 €				
SIN2	-46,02 €				
SIN3	-336,31 €				
SIN4	-908,05 €				

Tabela 9.3 - Rentabilidade Gerada - Análise Preliminar

9.1.1.7. Estimação da Probabilidade de Produzir Lucros

Características	Coeficientes	
(Intercept)	2,27E+00	β_0
DIAS_CONTRATO	-4,13E-04	β_1
CANALINTERNET	9,46E-02	β_2
CANALPARCEIROS	-6,08E-02	β_3
PRODUTOMOTO	9,17E-01	β_4
PRODUTOTOPPING	4,60E-01	β_5
FRACCIONAMENTOMENSAL	1,58E-01	β_6
FRACCIONAMENTOSEMESTRAL	-1,54E-01	β_7
FRACCIONAMENTOTRIMESTRAL	-1,14E-01	β_8

TEMPO_SEGUROTS2	-5,04E-01	β_9
TEMPO_SEGUROTS3	-4,56E-01	β_{10}
TEMPO_SEGUROTS4	-4,55E-01	β_{11}
TEMPO_SEGUROTS5	-4,47E-01	β_{12}
IDADE_VEICULOIV2	-1,64E-01	β_{13}
IDADE_VEICULOIV3	-3,66E-01	β_{14}
IDADE_VEICULOIV4	-3,44E-01	β_{15}
TIPO_VEICULOL	1,90E-01	β_{16}
TIPO_VEICULOM	6,70E-01	β_{17}
CILINDRADAC4	1,10E-01	β_{18}
COMBUSTIVELG	5,90E-02	β_{19}
COMBUSTIVELM	-1,06E+00	β_{20}
LUGARESL3	-3,62E-01	β_{21}
LUGARESL4	-1,46E-01	β_{22}
LUGARESL5	-2,72E-01	β_{23}
LUGARESL6	-4,34E-01	β_{24}
LUGARESL7	-2,39E-01	β_{25}
LUGARESL9	-2,77E-01	β_{26}
PREMIO_TOTAL	1,80E-04	β_{27}
SEXOM	5,62E-02	β_{28}
IDADEI3	1,15E-01	β_{29}
IDADEI7	4,57E-02	β_{30}
ANOS_CARTATC5	4,21E-01	β_{31}
ZONAZ3	1,32E-01	β_{32}
ZONAZ4	5,96E-02	β_{33}
ZONAZ5	1,99E-01	β_{34}
ZONAZ6	2,43E-01	β_{35}
ZONAZ7	4,96E-01	β_{36}
ZONAZ8	5,35E-01	β_{37}
ZONAZ9	6,07E-01	β_{38}
VSS	-1,06E-01	β_{39}
FRANQUIA2	4,72E-01	β_{40}
FRANQUIA4	3,39E-01	β_{41}
SIN_5_ANOSSIN2	-8,77E-01	β_{42}
SIN_5_ANOSSIN3	-1,48E+00	β_{43}
SIN_5_ANOSSIN4	-2,05E+00	β_{44}

Tabela 9.4 – Produção de Lucro - Coeficientes do Modelo Ajustado

Pela Equação (4.1.1) tem-se:

$$Y = \beta_0 + \beta_1 \text{DIAS_CONTRATO} + \beta_2 \text{CANALINTERNET} + \beta_3 \text{CANALPARCEIROS} + \beta_4 \text{PRODUTOMOTO} + \beta_5 \text{PRODUTOTOPPING} + \beta_6 \text{FRACCIONAMENTOMENSAL} + \beta_7 \text{FRACCIONAMENTO}$$

SEMESTRAL+ β 8FRACCIONAMENTOTRIMESTRAL+ β 9TEMPO_SEGUOTS2+ β 10TEMPO_SEGUOTS3+ β 11TEMPO_SEGUOTS4+ β 12TEMPO_SEGUOTS5+ β 13IDADE_VEICULOIV2+ β 14IDADE_VEICULOIV3+ β 15IDADE_VEICULOIV4+ β 16TIPO_VEICULOL+ β 17TIPO_VEICULOM+ β 18CILINDRADAC4+ β 19COMBUSTIVELG+ β 20COMBUSTIVELM+ β 21LUGARES3+ β 22LUGARES4+ β 23LUGARES5+ β 24LUGARES6+ β 25LUGARES7+ β 26LUGARES9+ β 27PREMIO_TOTAL+ β 28SEXOM+ β 29IDADEI3+ β 30IDADEI7+ β 31ANOS_CARTATC5+ β 32ZONA3+ β 33ZONA4+ β 34ZONA5+ β 35ZONA6+ β 36ZONA7+ β 37ZONA8+ β 38ZONA9+ β 39VSS+ β 40FRANQUIA2+ β 41FRANQUIA4+ β 42SIN_5_ANOSSIN2+ β 43SIN_5_ANOSSIN3+ β 44SIN_5_ANOSSIN4.

9.1.1.8. Estimação da Rentabilidade Gerada

Características	Coefficientes	
(Intercept)	1,07E+01	β 0
DIAS_CONTRATO	-9,42E-06	β 1
PRODUTOMOTO	1,58E-03	β 2
PRODUTOTOPPING	-1,31E-03	β 3
FRACCIONAMENTOMENSAL	-1,14E-03	β 4
FRACCIONAMENTOTRIMESTRAL	-3,51E-04	β 5
BONUS_MALUS	-6,05E-05	β 6
TEMPO_SEGUOTS2	1,34E-03	β 7
TEMPO_SEGUOTS3	1,48E-03	β 8
TEMPO_SEGUOTS4	1,21E-03	β 9
TEMPO_SEGUOTS5	1,36E-03	β 10
IDADE_VEICULOIV2	1,45E-03	β 11
IDADE_VEICULOIV3	3,11E-03	β 12
IDADE_VEICULOIV4	3,22E-03	β 13
CILINDRADAC3	-4,98E-04	β 14
CILINDRADAC4	-8,05E-04	β 15
COMBUSTIVELG	6,62E-04	β 16
LUGARES3	2,15E-03	β 17
LUGARES5	3,38E-04	β 18
PREMIO_TOTAL	8,08E-06	β 19
IDADEI2	-1,38E-03	β 20
IDADEI5	5,82E-04	β 21
ZONA3	-2,75E-04	β 22
ZONA7	-9,54E-04	β 23
ZONA8	-1,08E-03	β 24
QIVS	-1,17E-03	β 25
VSS	2,98E-04	β 26

FRANQUIA2	-4,72E-03	β_{27}
FRANQUIA4	-2,85E-03	β_{28}
AV_FNS	-2,68E-03	β_{29}
SIN_5_ANOSSIN2	4,47E-03	β_{30}
SIN_5_ANOSSIN3	1,17E-02	β_{31}
SIN_5_ANOSSIN4	2,25E-02	β_{32}

Tabela 9.5 – Rentabilidade Gerada -Coeficientes do Modelo Ajustado

Pela Equação (4.1.1) tem-se:

$Y = \beta_0 + \beta_1 \text{DIAS_CONTRATO} + \beta_2 \text{PRODUTOMOTO} + \beta_3 \text{PRODUTOTOPPING} + \beta_4 \text{FRACCI}$
 $\text{ONAMENTOMENSAL} + \beta_5 \text{FRACCIONAMENTOTRIMESTRAL} + \beta_6 \text{BONUS_MALUS} + \beta_7 \text{TEMPO_SEGUROTS2}$
 $+ \beta_8 \text{TEMPO_SEGUROTS3} + \beta_9 \text{TEMPO_SEGUROTS4} + \beta_{10} \text{TEMPO_SEGUROT}$
 $\text{S5} + \beta_{11} \text{IDADE_VEICULOIV2} + \beta_{12} \text{IDADE_VEICULOIV3} + \beta_{13} \text{IDADE_VEICULOIV4} + \beta_{14} \text{CILIN}$
 $\text{DRADAC3} + \beta_{15} \text{CILINDRADAC4} + \beta_{16} \text{COMBUSTIVELG} + \beta_{17} \text{LUGARES3} + \beta_{18} \text{LUGARES5} + \beta_{19}$
 $\text{PREMIO_TOTAL} + \beta_{20} \text{IDADEI2} + \beta_{21} \text{IDADEI5} + \beta_{22} \text{ZONAZ3} + \beta_{23} \text{ZONAZ7} + \beta_{24} \text{ZONAZ8} + \beta_{25}$
 $\text{QIVS} + \beta_{26} \text{VSS} + \beta_{27} \text{FRANQUIA2} + \beta_{28} \text{FRANQUIA4} + \beta_{29} \text{AV_FNS} + \beta_{30} \text{SIN_5_ANOSSIN2} +$
 $\beta_{31} \text{SIN_5_ANOSSIN3} + \beta_{32} \text{SIN_5_ANOSSIN4}.$

9.2. INPUTS E OUTPUTS DO R

9.2.1. Declaração de Variáveis

```
> sample<-read.table(sampleFile, na.strings = c("NaN", ""), header=TRUE, sep="\t", dec=",")
> cat("Nomes da coluna: ", names(sample), "\n")
Nomes da coluna: NUMERO_APOLICE DIAS_CONTRATO CANAL PRODUTO FRACCIONAMENTO
BONUS_MALUS TEMPO_SEGURO IDADE_VEICULO TIPO_VEICULO CILINDRADA COMBUSTIVEL LUGARES
PREMIO_TOTAL SEXO IDADE ANOS_CARTA ZONA QIV VS FRANQUIA AV_FN IE SIN_5_ANOS R
RENTABILIDADE RENT_POSIT
> sampleSize=dim(sample)[1]
> names(sample)
[1] "NUMERO_APOLICE" "DIAS_CONTRATO" "CANAL"      "PRODUTO"
[5] "FRACCIONAMENTO" "BONUS_MALUS"  "TEMPO_SEGURO" "IDADE_VEICULO"
[9] "TIPO_VEICULO"  "CILINDRADA"   "COMBUSTIVEL"  "LUGARES"
[13] "PREMIO_TOTAL" "SEXO"         "IDADE"        "ANOS_CARTA"
[17] "ZONA"          "QIV"          "VS"           "FRANQUIA"
[21] "AV_FN"         "IE"           "SIN_5_ANOS"
[25] "R"             "RENTABILIDADE" "RENT_POSIT"
> TEMPO_SEGURO=factor(sample$TEMPO_SEGURO)
> IDADE_VEICULO=factor(sample$IDADE_VEICULO)
> CILINDRADA=factor(sample$CILINDRADA)
> LUGARES=factor(sample$LUGARES)
> IDADE=factor(sample$IDADE)
> ANOS_CARTA=factor(sample$ANOS_CARTA)
> ZONA=factor(sample$ZONA)
> FRANQUIA=factor(sample$FRANQUIA)
> SIN_5_ANOS=factor(sample$SIN_5_ANOS)
> R=factor(sample$R)
```

9.2.2. Análise Estatística das variáveis

```
summary(sample$R)
```

```
Min.    1st Qu.  Median Mean    3rd Qu.  Max.
0.0000  1.0000  1.0000  0.9033  1.0000  1.0000
```

```
summary(sample$RENTABILIDADE)
```

```
Min.      1st Qu.  Median Mean    3rd Qu.  Max.
-44400.00  75.48  196.70  128.50  403.00  5648.00
```

```
summary(sample$RENT_POSIT)
```

```
Min.    1st Qu.  Median Mean    3rd Qu.  Max.
38750   44000  44200  44270  44330  88800
```

9.2.3. Produção de Lucro

9.2.3.1. Produção de Lucro - Resultados

Modelo Nulo

```
#modelo nulo--> lucronull
> summary(lucronull)
Call: glm(formula = sample$R ~ 1, family = binomial())
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1616	0.4509	0.4509	0.4509	0.4509

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.234699	0.007761	288	<2e-16 ***

AIC: 120817

Modelo completo

```
#modelo completo--> lucrofull
```

> summary(lucrofull)

```
Call: glm(formula = sample$R ~ sample$DIAS_CONTRATO + sample$CANAL +
  sample$PRODUTO + sample$FRACCIONAMIENTO + +sample$BONUS_MALUS +
  TEMPO_SEGURO + IDADE_VEICULO + sample$TIPO_VEICULO + CILINDRADA +
  sample$COMBUSTIVEL + +LUGARES + sample$PREMIO_TOTAL + sample$SEXO +
  IDADE + ANOS_CARTA + ZONA + sample$QIV + sample$VS + FRANQUIA +
  sample$AV_FN + sample$IE + SIN_5_ANOS, family = binomial())
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9849	0.3376	0.4221	0.4804	1.4522

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2,19E+00	5,74E-01	3,81E+00	0.000137	***
sample\$DIAS_CONTRATO	-4,10E-04	2,08E-05	-1,97E+01	< 2e-16	***
sample\$CANALINTERNET	1,05E-01	2,19E-02	4,80E+00	1.57e-06	***
sample\$CANALPARCEIROS	-5,92E-02	2,04E-02	-2,90E+00	0.003710	**
sample\$PRODUTOLIGHT	2,47E-01	2,19E-01	1,13E+00	0.259076	
sample\$PRODUTOMAX	-3,12E-01	2,22E-01	-1,40E+00	0.160460	
sample\$PRODUTOMOTO	7,47E-01	2,88E-01	2,60E+00	0.009469	**
sample\$PRODUTOSAFE	1,42E-01	2,19E-01	6,48E-01	0.517075	
sample\$PRODUTOTOPPING	2,64E-01	2,18E-01	1,21E+00	0.225648	
sample\$FRACCIONAMENTOMENSAL	1,48E-01	4,41E-02	3,35E+00	0.000798	***
sample\$FRACCIONAMENTOSEMESTRA	-1,56E-01	1,95E-02	-7,97E+00	1.65e-15	***
sample\$FRACCIONAMENTOTRIMESTR	-1,21E-01	2,35E-02	-5,14E+00	2.74e-07	***
sample\$BONUS_MALUS	-1,51E-03	1,03E-03	-1,47E+00	0.142007	
TEMPO_SEGUROS2	-4,91E-01	8,87E-02	-5,54E+00	3.07e-08	***
TEMPO_SEGUROS3	-4,55E-01	9,17E-02	-4,97E+00	6.82e-07	***
TEMPO_SEGUROS4	-4,65E-01	9,09E-02	-5,12E+00	3.03e-07	***
TEMPO_SEGUROS5	-4,45E-01	8,60E-02	-5,18E+00	2.23e-07	***
IDADE_VEICULOIV2	-1,68E-01	6,76E-02	-2,48E+00	0.013179	*
IDADE_VEICULOIV3	-3,74E-01	6,96E-02	-5,38E+00	7.61e-08	***
IDADE_VEICULOIV4	-3,51E-01	7,32E-02	-4,79E+00	1.68e-06	***
sample\$TIPO_VEICULOC	-1,57E-01	8,21E-02	-1,91E+00	0.056329	.
sample\$TIPO_VEICULOL	6,37E-02	5,84E-02	1,09E+00	0.275713	
sample\$TIPO_VEICULOM	5,61E-01	1,92E-01	2,92E+00	0.003490	**
CILINDRADAC2	5,45E-02	2,62E-02	2,08E+00	0.037678	*
CILINDRADAC3	5,84E-02	3,10E-02	1,88E+00	0.060041	.
CILINDRADAC4	1,67E-01	5,61E-02	2,98E+00	0.002930	**

CILINDRADAC5	7,59E-02	7,88E-02	9,62E-01	0.335930	
sample\$COMBUSTIVELE	7,59E+00	3,66E+01	2,07E-01	0.835645	
sample\$COMBUSTIVELG	8,61E-02	2,38E-02	3,62E+00	0.000297	***
sample\$COMBUSTIVELM	-1,05E+00	5,26E-01	-1,99E+00	0.047033	*
LUGARES2	-9,10E-02	1,00E-01	-9,07E-01	0.364324	
LUGARES3	-4,55E-01	1,33E-01	-3,42E+00	0.000625	***
LUGARES4	-2,62E-01	1,20E-01	-2,18E+00	0.029289	*
LUGARES5	-3,96E-01	1,14E-01	-3,47E+00	0.000524	***
LUGARES6	-5,40E-01	1,71E-01	-3,16E+00	0.001572	**
LUGARES7	-3,68E-01	1,27E-01	-2,90E+00	0.003697	**
LUGARES8	-1,25E-01	2,73E-01	4,58E-01	0.646688	
LUGARES9	-3,86E-01	1,66E-01	-2,33E+00	0.020035	*
sample\$PREMIO_TOTAL	4,95E-04	1,33E-04	3,71E+00	0.000205	***
sample\$SEXOM	4,61E-02	1,82E-02	2,53E+00	0.011314	*
IDADEI2	3,63E-02	1,38E-01	2,63E-01	0.792459	
IDADEI3	9,25E-02	1,24E-01	7,49E-01	0.453975	
IDADEI4	9,18E-02	1,24E-01	7,38E-01	0.460295	
IDADEI5	9,97E-02	1,25E-01	8,00E-01	0.423485	
IDADEI6	-1,39E-02	1,25E-01	-1,11E-01	0.911696	
IDADEI7	1,09E-02	1,26E-01	8,60E-02	0.931236	
IDADEI8	5,54E-02	1,29E-01	4,30E-01	0.667220	
ANOS_CARTATC2	2,19E-01	4,84E-01	4,52E-01	0.651059	
ANOS_CARTATC3	4,45E-01	4,84E-01	9,20E-01	0.357770	
ANOS_CARTATC4	6,68E-01	4,84E-01	1,38E+00	0.167571	
ANOS_CARTATC5	9,26E-01	4,84E-01	1,91E+00	0.055924	.
ZONA2	1,45E-02	3,35E-02	4,34E-01	0.664490	
ZONA3	1,43E-01	3,28E-02	4,38E+00	1.21e-05	***
ZONA4	7,41E-02	3,47E-02	2,14E+00	0.032655	*
ZONA5	2,18E-01	3,44E-02	6,32E+00	2.55e-10	***
ZONA6	2,64E-01	3,80E-02	6,95E+00	3.65e-12	***
ZONA7	5,11E-01	4,83E-02	1,06E+01	< 2e-16	***
ZONA8	5,54E-01	7,43E-02	7,46E+00	8.98e-14	***
ZONA9	6,25E-01	7,83E-02	7,99E+00	1.40e-15	***
sample\$QIVS	1,27E-02	2,19E-02	5,82E-01	0.560834	
sample\$VSS	-1,14E-01	2,31E-02	-4,93E+00	8.28e-07	***
FRANQUIA2	5,12E-01	5,55E-02	9,23E+00	< 2e-16	***
FRANQUIA4	4,08E-01	6,19E-02	6,59E+00	4.48e-11	***
sample\$AV_FNS	1,04E-02	6,18E-02	1,67E-01	0.867006	
sample\$IES	6,17E-03	2,92E-02	2,11E-01	0.832516	
SIN_5_ANOSSIN2	-8,86E-01	2,66E-02	-3,33E+01	< 2e-16	***
SIN_5_ANOSSIN3	-1,49E+00	6,06E-02	-2,45E+01	< 2e-16	***
SIN_5_ANOSSIN4	-2,05E+00	1,09E-01	-1,89E+01	< 2e-16	***

AIC: 115854

9.2.3.2. Comparação dos dois modelos (modelo nulo e modelo completo)

```
> anova(lucronull,lucrofull,test="Chisq")
```

Analysis of Deviance Table

Model 1: sample\$R ~ 1

Model 2: sample\$R ~ sample\$DIAS_CONTRATO + sample\$CANAL + sample\$PRODUTO + sample\$FRACCIONAMENTO + +sample\$BONUS_MALUS + TEMPO_SEGURO + IDADE_VEICULO + sample\$TIPO_VEICULO + CILINDRADA + sample\$COMBUSTIVEL + +LUGARES + sample\$PREMIO_TOTAL + sample\$SEXO + IDADE + ANOS_CARTA + ZONA + sample\$QIV + sample\$VS + FRANQUIA + sample\$AV_FN + +sample\$IE + SIN_5_ANOS

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190126	120815			
2	190059	115718	67	5096.4	< 2.2e-16 ***

9.2.3.3. Seleção do modelo através do Método Stepwise-Backward

Variável	Pr(>Chi)	Resultado
IDADE	1.707e-05 ***	MANTER
AV_FN	0.8671	RETIRAR
COMBUSTIVEL	0.0005161 ***	MANTER
IE	0.8224	RETIRAR
ZONA	< 2.2e-16 ***	MANTER
ANOS_CARTA	< 2.2e-16 ***	MANTER
LUGARES	2.015e-09 ***	MANTER
QIV	0.5674	RETIRAR
PRODUTO	< 2.2e-16 ***	MANTER
CILINDRADA	0.04412 *	MANTER
TIPO_VEICULO	3.289e-05 ***	MANTER
BM	0.1302	RETIRAR

9.2.3.4. Modelo Agrupado (LUCRO_AGRUP)

```
> summary(LUCRO_AGRUP)
```

Call: glm(formula = sample\$R ~ sample\$DIAS_CONTRATO + sample\$CANAL + sample\$PRODUTO + sample\$FRACCIONAMENTO + +TEMPO_SEGURO + IDADE_VEICULO + sample\$TIPO_VEICULO + CILINDRADA + sample\$COMBUSTIVEL + +LUGARES + sample\$PREMIO_TOTAL + sample\$SEXO + IDADE + ANOS_CARTA + ZONA + sample\$VS + FRANQUIA + +SIN_5_ANOS, family = binomial())

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9842	0.3373	0.4234	0.4822	1.4157

Coefficients:

	Estimate	Std.Error	zvalue	Pr(> z)	
(Intercept)	2.27E+00	1.36E-01	16.774	<2e-16	***
sample\$DIAS_CONTRATO	-4.13E-04	2.06E-05	-20.095	<2e-16	***
sample\$CANALINTERNET	9.46E-02	2.11E-02	4.478	7.52e-06	***
sample\$CANALPARCEIROS	-6.08E-02	2.02E-02	-3.016	0.002564	**
sample\$PRODUTOMOTO	9.17E-01	1.93E-01	4.758	1.96e-06	***

sample\$PRODUTOTOPPING	4.60E-01	4.20E-02	10.949	<2e-16	***
sample\$FRACCIONAMENTOMENSAL	1.58E-01	4.35E-02	3.619	0.000296	***
sample\$FRACCIONAMENTOSEMESTRAL	-1.54E-01	1.95E-02	-7.926	2.26e-15	***
sample\$FRACCIONAMENTOTRIMESTRAL	-1.14E-01	2.33E-02	-4.877	1.08e-06	***
TEMPO_SEGUROS2	-5.04E-01	8.85E-02	-5.702	1.19e-08	***
TEMPO_SEGUROS3	-4.56E-01	9.14E-02	-4.987	6.14e-07	***
TEMPO_SEGUROS4	-4.55E-01	9.07E-02	-5.017	5.24e-07	***
TEMPO_SEGUROS5	-4.47E-01	8.58E-02	-5.209	1.90e-07	***
IDADE_VEICULOIV2	-1.64E-01	6.70E-02	-2.445	0.014488	*
IDADE_VEICULOIV3	-3.66E-01	6.90E-02	-5.294	1.20e-07	***
IDADE_VEICULOIV4	-3.44E-01	7.25E-02	-4.742	2.12e-06	***
sample\$TIPO_VEICULOL	1.90E-01	5.99E-02	3.166	0.001547	**
sample\$TIPO_VEICULOM	6.70E-01	1.87E-01	3.586	0.000336	***
CILINDRADAC4	1.10E-01	4.68E-02	2.346	0.018964	*
sample\$COMBUSTIVELG	5.90E-02	2.14E-02	2.755	0.005866	**
sample\$COMBUSTIVELM	-1.06E+00	5.26E-01	-2.014	0.043963	*
LUGARES3	-3.62E-01	8.69E-02	-4.169	3.05e-05	***
LUGARES4	-1.46E-01	6.58E-02	-2.212	0.026953	*
LUGARES5	-2.72E-01	5.40E-02	-5.040	4.67e-07	***
LUGARES6	-4.34E-01	1.38E-01	-3.145	0.001661	**
LUGARES7	-2.39E-01	7.62E-02	-3.130	0.001745	**
LUGARES9	-2.77E-01	1.32E-01	-2.108	0.034989	*
sample\$PREMIO_TOTAL	1.80E-04	1.09E-04	1.658	0.097320	.
sample\$SEXOM	5.62E-02	1.80E-02	3.129	0.001753	**
IDADEI3	1.15E-01	1.94E-02	5.923	3.15e-09	***
IDADEI7	4.57E-02	2.55E-02	1.791	0.073266	.
ANOS_CARTATC5	4.21E-01	2.77E-02	15.167	<2e-16	***
ZONA3	1.32E-01	2.25E-02	5.854	4.80e-09	***
ZONA4	5.96E-02	2.51E-02	2.375	0.017541	*
ZONA5	1.99E-01	2.45E-02	8.119	4.70e-16	***
ZONA6	2.43E-01	2.93E-02	8.302	<2e-16	***
ZONA7	4.96E-01	4.18E-02	11.868	<2e-16	***
ZONA8	5.35E-01	7.01E-02	7.638	2.21e-14	***
ZONA9	6.07E-01	7.43E-02	8.175	2.95e-16	***
sample\$VSS	-1.06E-01	2.09E-02	-5.055	4.31e-07	***
FRANQUIA2	4.72E-01	5.46E-02	8.641	<2e-16	***
FRANQUIA4	3.39E-01	5.94E-02	5.712	1.12e-08	***
SIN_5_ANOSSIN2	-8.77E-01	2.65E-02	-33.132	<2e-16	***
SIN_5_ANOSSIN3	-1.48E+00	6.03E-02	-24.511	<2e-16	***
SIN_5_ANOSSIN4	-2.05E+00	1.09E-01	-18.829	<2e-16	***

AIC: 115902

9.2.4. Rentabilidade Gerada

9.2.4.1. Rentabilidade Gerada - Resultados

Modelo Nulo

```
#modelo nulo--> rentnull
```

```
rentnull=glm(sample$RENT_POSIT~1,family=Gamma)
```

```
summary(rentnull)
```

```
Call: glm(formula = sample$RENT_POSIT ~ 1, family = Gamma)
```

Deviance Residuals:

```
Min      1Q      Median      3Q      Max
-0.13027 -0.00621 -0.00154  0.00120  0.78710
```

Coefficients:

```
              Estimate      Std. Error  t value  Pr(>|t|)
(Intercept)  2.259e-05   1.294e-09  17460   <2e-16 ***
```

```
AIC: 3170955
```

Modelo Completo

```
#modelo completo--> rentfull
```

```
> summary(rentfull)
```

```
Call: glm(formula = sample$R ~ sample$DIAS_CONTRATO + sample$CANAL +
  sample$PRODUTO + sample$FRACCIONAMIENTO + +sample$BONUS_MALUS +
  TEMPO_SEGURO + IDADE_VEICULO + sample$TIPO_VEICULO + CILINDRADA +
  sample$COMBUSTIVEL + +LUGARES + sample$PREMIO_TOTAL + sample$SEXO +
  IDADE + ANOS_CARTA + ZONA + sample$QIV + sample$VS + FRANQUIA +
  sample$AV_FN +sample$IE + SIN_5_ANOS, family = Gamma(link = log))
```

Deviance Residuals:

```
Min      1Q      Median      3Q      Max
-0.12264 -0.00464 -0.00315 -0.00143  0.78628
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1,07E+01	4,38E-03	2,44E+03	< 2e-16	***
sample\$DIAS_CONTRATO	-9,44E-06	1,51E-07	-6,24E+01	< 2e-16	***
sample\$CANALINTERNET	1,10E-04	1,53E-04	7,14E-01	0.474931	
sample\$CANALPARCEIROS	5,28E-05	1,49E-04	3,54E-01	0.723476	
sample\$PRODUTOLIGHT	3,34E-03	1,64E-03	2,04E+00	0.041179	*
sample\$PRODUTOMAX	1,63E-03	1,67E-03	9,74E-01	0.330107	
sample\$PRODUTOMOTO	5,99E-03	1,99E-03	3,02E+00	0.002538	**
sample\$PRODUTOSAFE	1,20E-03	1,64E-03	7,33E-01	0.463500	
sample\$PRODUTOTOPPING	2,13E-03	1,63E-03	1,30E+00	0.192100	
sample\$FRACCIONAMENTOMENSAL	-1,25E-03	3,08E-04	-4,07E+00	4.77e-05	***
sample\$FRACCIONAMENTOSEMESTRAL	-1,78E-04	1,40E-04	-1,27E+00	0.203361	
sample\$FRACCIONAMENTOTRIMESTRAL	-4,96E-04	1,70E-04	-2,92E+00	0.003452	**
sample\$BONUS_MALUS	-6,43E-05	7,16E-06	-8,97E+00	< 2e-16	***
TEMPO_SEGUROS2	1,44E-03	4,89E-04	2,95E+00	0.003160	**
TEMPO_SEGUROS3	1,56E-03	5,17E-04	3,03E+00	0.002488	**
TEMPO_SEGUROS4	1,31E-03	5,08E-04	2,58E+00	0.009958	**
TEMPO_SEGUROS5	1,44E-03	4,62E-04	3,11E+00	0.001896	**

IDADE_VEICULOIV2	1,43E-03	4,54E-04	3,15E+00	0.001615	**
IDADE_VEICULOIV3	3,02E-03	4,64E-04	6,52E+00	7.24e-11	***
IDADE_VEICULOIV4	3,14E-03	4,89E-04	6,41E+00	1.47e-10	***
sample\$TIPO_VEICULOC	9,29E-04	5,82E-04	1,60E+00	0.110374	
sample\$TIPO_VEICULOL	1,05E-03	4,25E-04	2,47E+00	0.013418	*
sample\$TIPO_VEICULOM	-2,81E-04	1,18E-03	-2,37E-01	0.812462	
CILINDRADAC2	-3,09E-04	1,89E-04	-1,64E+00	0.101884	
CILINDRADAC3	-7,17E-04	2,25E-04	-3,20E+00	0.001401	**
CILINDRADAC4	-9,48E-04	3,99E-04	-2,38E+00	0.017416	*
CILINDRADAC5	2,02E-04	5,64E-04	3,59E-01	0.719601	
sample\$COMBUSTIVELE	-4,22E-03	7,76E-03	-5,44E-01	0.586644	
sample\$COMBUSTIVELG	6,95E-04	1,70E-04	4,08E+00	4.54e-05	***
sample\$COMBUSTIVELM	9,76E-03	5,35E-03	1,82E+00	0.068078	.
LUGARES2	6,47E-04	4,18E-04	1,55E+00	0.121325	
LUGARES3	2,62E-03	7,82E-04	3,35E+00	0.000819	***
LUGARES4	1,58E-04	6,17E-04	2,56E-01	0.798044	
LUGARES5	7,41E-04	5,58E-04	1,33E+00	0.184574	
LUGARES6	1,23E-03	1,16E-03	1,06E+00	0.289963	
LUGARES7	2,09E-04	6,88E-04	3,04E-01	0.761349	
LUGARES8	-1,60E-04	1,68E-03	-9,50E-02	0.924115	
LUGARES9	1,49E-03	1,07E-03	1,40E+00	0.163015	
sample\$PREMIO_TOTAL	9,14E-06	9,64E-07	9,48E+00	< 2e-16	***
sample\$SEXOM	-1,47E-04	1,35E-04	-1,09E+00	0.276414	
IDADEI2	-1,50E-03	1,11E-03	-1,35E+00	0.176422	
IDADEI3	-2,17E-04	9,96E-04	-2,18E-01	0.827317	
IDADEI4	6,02E-04	1,00E-03	6,02E-01	0.547449	
IDADEI5	9,75E-04	1,00E-03	9,73E-01	0.330661	
IDADEI6	1,04E-03	1,01E-03	1,04E+00	0.299403	
IDADEI7	8,48E-04	1,01E-03	8,36E-01	0.402986	
IDADEI8	3,21E-04	1,03E-03	3,11E-01	0.756064	
ANOS_CARTATC2	-4,19E-05	3,80E-03	-1,10E-02	0.991213	
ANOS_CARTATC3	9,69E-06	3,80E-03	3,00E-03	0.997963	
ANOS_CARTATC4	-4,29E-04	3,80E-03	-1,13E-01	0.910122	
ANOS_CARTATC5	-4,11E-04	3,80E-03	-1,08E-01	0.913785	
ZONA2	-1,56E-04	2,54E-04	-6,14E-01	0.539240	
ZONA3	-4,87E-04	2,46E-04	-1,98E+00	0.047448	*
ZONA4	-1,81E-04	2,62E-04	-6,91E-01	0.489880	
ZONA5	-3,55E-04	2,54E-04	-1,40E+00	0.161406	
ZONA6	-8,79E-05	2,77E-04	-3,18E-01	0.750560	
ZONA7	-1,14E-03	3,20E-04	-3,57E+00	0.000361	***
ZONA8	-1,28E-03	4,58E-04	-2,80E+00	0.005176	**
ZONA9	-4,64E-04	4,75E-04	-9,78E-01	0.328290	
sample\$QIVS	-1,37E-03	1,62E-04	-8,47E+00	< 2e-16	***
sample\$VSS	3,97E-04	1,75E-04	2,26E+00	0.023679	*

FRANQUIA2	-4,43E-03	3,87E-04	-1,15E+01	< 2e-16	***
FRANQUIA4	-2,40E-03	4,33E-04	-5,54E+00	3.00e-08	***
sample\$AV_FNS	-2,78E-03	4,92E-04	-5,65E+00	1.58e-08	***
sample\$IES	-8,76E-07	2,22E-04	-4,00E-03	0.996851	
SIN_5_ANOSSIN2	4,46E-03	2,47E-04	1,80E+01	< 2e-16	***
SIN_5_ANOSSIN3	1,16E-02	6,74E-04	1,73E+01	< 2e-16	***
SIN_5_ANOSSIN4	2,26E-02	1,29E-03	1,75E+01	< 2e-16	***

AIC: 3162926

9.2.4.2. Comparação dos dois modelos

```
> anova(rentnull.rentfull.test="Chisq")
```

Analysis of Deviance Table

Model 1: sample\$RENT_POSIT ~ 1

Model 2: sample\$RENT_POSIT ~ sample\$DIAS_CONTRATO + sample\$CANAL + sample\$PRODUTO + sample\$FRACCIONAMENTO + +sample\$BONUS_MALUS + TEMPO_SEGURO + IDADE_VEICULO + sample\$TIPO_VEICULO + CILINDRADA + sample\$COMBUSTIVEL + +LUGARES + sample\$PREMIO_TOTAL + sample\$SEXO + IDADE + ANOS_CARTA + ZONA + sample\$QIV + sample\$VS + FRANQUIA + sample\$AV_FN + +sample\$IE + SIN_5_ANOS

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	190126	99.468			
2	190059	95.288	67	4.1798	< 2.2e-16 ***

9.2.4.3. Seleção do modelo através do Método Stepwise-Backward

Variavel	Pr(>Chi)	Resultado
ANOS_CARTA	0.797	RETIRAR
IE	0.9971	RETIRAR
LUGARES	0.02162 *	MANTER
IDADE	1.057e-05 ***	MANTER
TIPO_VEICULO	0.05676 .	RETIRAR
CILINDRADA	0.004254 **	MANTER
ZONA	0.001096 **	MANTER
CANAL	0.8034	RETIRAR
COMBUSTIVEL	4.593e-05 ***	MANTER
PRODUTO	< 2.2e-16 ***	MANTER
SEXO	0.2806	RETIRAR
FRACCIONAMENTO	0.0001236 ***	MANTER

9.2.4.4. Modelo Agrupado (RENT_AGRUP)

```
> summary(rent_agrup)
```

Call: glm(sample\$RENT_POSIT~sample\$DIAS_CONTRATO+sample\$PRODUTO+sample\$FRACCIONAMENTO+sample\$BONUS_MALUS+TEMPO_SEGURO+IDADE_VEICULO+CILINDRADA+sample\$CO

MBUSTIVEL+LUGARES+sample\$PREMIO_TOTAL+IDADE+ZONA+sample\$QIV+sample\$VSS+FRANQUIA+sample\$AV_FN+SIN_5_ANOS,family = Gamma(link = log))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.12259	-0.00463	-0.00312	-0.00144	0.78653

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1,07E+01	7,43E-04	1,44E+04	< 2e-16	***
sample\$DIAS_CONTRATO	-9,42E-06	1,46E-07	-6,45E+01	< 2e-16	***
sample\$PRODUTOMOTO	1,58E-03	2,91E-04	5,44E+00	5,26E-08	***
sample\$PRODUTOTOPPING	-1,31E-03	1,77E-04	-7,39E+00	1,46E-13	***
sample\$FRACCIONAMENTOMENSAL	-1,14E-03	2,94E-04	-3,87E+00	0,000109	***
sample\$FRACCIONAMENTOTRIMESTRAL	-3,51E-04	1,47E-04	-2,40E+00	0,016631	*
sample\$BONUS_MALUS	-6,05E-05	5,70E-06	-1,06E+01	< 2e-16	***
TEMPO_SEGUROS2	1,34E-03	4,87E-04	2,76E+00	0,00583	**
TEMPO_SEGUROS3	1,48E-03	5,15E-04	2,88E+00	0,004009	**
TEMPO_SEGUROS4	1,21E-03	5,07E-04	2,39E+00	0,017037	*
TEMPO_SEGUROS5	1,36E-03	4,61E-04	2,95E+00	0,003201	**
IDADE_VEICULOIV2	1,45E-03	4,49E-04	3,23E+00	0,001256	**
IDADE_VEICULOIV3	3,11E-03	4,56E-04	6,82E+00	9,40E-12	***
IDADE_VEICULOIV4	3,22E-03	4,81E-04	6,69E+00	2,18E-11	***
CILINDRADAC3	-4,98E-04	1,53E-04	-3,25E+00	0,00116	**
CILINDRADAC4	-8,05E-04	3,27E-04	-2,46E+00	0,0138	*
sample\$COMBUSTIVELG	6,62E-04	1,60E-04	4,13E+00	3,63E-05	***
LUGARES3	2,15E-03	6,23E-04	3,46E+00	0,000549	***
LUGARES5	3,38E-04	1,73E-04	1,96E+00	0,050069	.
sample\$PREMIO_TOTAL	8,08E-06	8,09E-07	9,98E+00	< 2e-16	***
IDADEI2	-1,38E-03	5,90E-04	-2,34E+00	0,019248	*
IDADEI5	5,82E-04	1,36E-04	4,30E+00	1,73E-05	***
ZONAZ3	-2,75E-04	1,37E-04	-2,01E+00	0,04452	*
ZONAZ7	-9,54E-04	2,42E-04	-3,94E+00	8,21E-05	***
ZONAZ8	-1,08E-03	4,06E-04	-2,65E+00	0,008041	**
sample\$QIVS	-1,17E-03	1,50E-04	-7,80E+00	6,41E-15	***
sample\$VSS	2,98E-04	1,63E-04	1,82E+00	0,068599	.
FRANQUIA2	-4,72E-03	3,72E-04	-1,27E+01	< 2e-16	***
FRANQUIA4	-2,85E-03	4,07E-04	-7,02E+00	2,29E-12	***
sample\$AV_FNS	-2,68E-03	3,28E-04	-8,16E+00	3,48E-16	***
SIN_5_ANOSSIN2	4,47E-03	2,46E-04	1,82E+01	< 2e-16	***
SIN_5_ANOSSIN3	1,17E-02	6,72E-04	1,73E+01	< 2e-16	***
SIN_5_ANOSSIN4	2,25E-02	1,29E-03	1,75E+01	< 2e-16	***

AIC: 3162917

9.3. GLOSSÁRIO

Apólice de seguro: Documento que contém as condições do contrato de seguro acordadas pelas partes e que incluem as condições gerais, especiais e particulares.

Automóvel Ligeiro: Veículo com peso bruto igual ou inferior a 3500kg e com lotação não superior a nove lugares, incluindo o do condutor.

Capital Seguro: valor máximo que o segurador paga em caso de sinistro, mesmo que o prejuízo seja superior. Este valor é normalmente definido nas condições particulares da apólice.

Certificado de seguro/Carta verde: Documento que confirma que um contrato de seguro é válido. Pode ser entregue pelo segurador ou por um mediador de seguros.

Ciclomotor: Veículo dotado de duas ou três rodas, com velocidade máxima, em patamar e por construção, não superior a 45km/h. e cujo motor não tenha cilindrada superior a 50cm³.

Coberturas: Conjunto de situações cuja verificação determina a prestação do segurador ao abrigo do contrato.

Condições Especiais: Conjunto de cláusulas que se destina a esclarecer, completar ou alterar disposições das condições gerais.

Condições Gerais: Conjunto de cláusulas oficialmente aprovadas que define o regula todos os aspectos genéricos inerentes a cada ramo ou tipo de seguro e que tornam o contrato uniforme, bem como os direitos e obrigação de cada uma das partes.

Condições Particulares: Conjuntos de cláusulas que identifica o tomador de seguro, o objecto de seguro enunciando os restantes elementos próprios de cada contrato. Das condições particulares constam a empresa de seguros, tomador, âmbito do contrato, garantias, riscos cobertos e exclusões, definição das opções, início e duração do contrato.

Dano corporal: Dano relativo à vida. à saúde ou integridade física de uma pessoa.

Dano material: Prejuízo causado a coisas, bens materiais, créditos e qualquer outro direito patrimonial.

Dano: Prejuízo sofrido por alguém. O dano pode ser causado por perda, destruição ou avaria de bens ou por lesão de que afecte a saúde física ou mental de uma pessoa.

Franquia: Parte do valor dos danos que fica a cargo do tomador do seguro ou segurado.

Fundo de Garantia Automóvel: Entidade que garante o pagamento das indemnizações por danos corporais e materiais resultantes de acidentes de viação causados por veículos matriculados em Portugal que não tenham, à data do acidente, o seguro obrigatório necessário.

Mediador de seguros: Qualquer pessoa ou entidade que exerça, mediante remuneração, a actividade de mediação de seguros e se encontra inscrito como mediador no Instituto de Seguros de Portugal (ISP). Pode fazê-lo por conta de um ou vários seguradores, ou de forma independente.

Motociclos: Veículo dotado de duas rodas, com ou sem carro lateral, com motor de propulsão com cilindrada superior a 50cm³, no caso de motor a combustão interna, ou que, por construção, exceda em patamar a velocidade de 45km/h.

Período de carência: Período entre o início do contrato de seguro e uma determinada data, no qual certas coberturas não se encontram ainda a produzir efeitos.

Prémio: Valor total, incluindo taxas e impostos, que o tomador do seguro deve pagar ao segurador pelo seguro.

Proposta de seguro ou Simulação: Documento através do qual o tomador do seguro expressa a vontade de celebrar o contrato de seguro e dá a conhecer ao segurador o risco que pretende segurar.

Risco: Incerteza associada a um acontecimento futuro, seja quanto à sua realização, ao momento em que ocorreu ou aos danos dele decorrentes.

Segurado: Pessoa ou entidade no interesse da qual é feito o contrato de seguro ou pessoa cuja vida, saúde ou integridade física se segura (pessoa segura).

Segurador: Entidade legalmente autorizada a exercer a actividade seguradora e que é parte no contrato de seguro.

Seguro de danos próprios: Designação por que é conhecido o seguro que cobre os prejuízos sofridos pelo veículo seguro, mesmo nas situações em que o condutor seja responsável pelo acidente.

Sinistro: Evento ou serie de eventos que resultam de uma mesma causa e que accionam a cobertura do risco prevista no contrato.

Tomador de seguro: Pessoa que celebra o contrato de seguro com a empresa de seguros. sendo responsável pelo pagamento do prémio.

Valor do salvado: valor do veículo no estado em que ficou após o acidente.

Valor venal: valor pelo qual o veículo poderia ser substituído imediatamente antes do acidente.