



Universidade Eduardo Mondlane

Faculdade de Ciências

Departamento de Matemática e Informática

TRABALHO DE LICENCIATURA

**MODELOS PARA ANÁLISE DE RISCO DE CRÉDITO:
APLICAÇÃO NUMA INSTITUIÇÃO FINANCEIRA**

Autora: Sheyla Ratan Rodrigues Cassy

Maputo, Fevereiro de 2009



Universidade Eduardo Mondlane
Faculdade de Ciências
Departamento de Matemática e Informática

TRABALHO DE LICENCIATURA

**MODELOS PARA ANÁLISE DE RISCO DE CRÉDITO:
APLICAÇÃO NUMA INSTITUIÇÃO FINANCEIRA**

Autora: Sheyla Ratan Rodrigues Cassy

Supervisora: Dra. Rafica Abdul Razac

**"A ciência não pode prever o que vai acontecer.
Só pode prever a probabilidade de algo acontecer."**

(César Lattes)

À memória da minha avó **Tereza Rodrigues** que muito fez por mim,
À minha pequenina **Gysella Rodrigues** que me tem como espelho,
Aos meus pais **Bhangy Cassy** e **Alexandra Rodrigues**,
Toda a **Família Rodrigues** que acompanha cada passo que dou.

Agradecimentos

- À Dra. **Rafica Abdul Razac** pela sua disponibilidade, por ter acreditado neste projecto desde o início e pela paciência durante elaboração deste trabalho, pois várias foram as dificuldades que encontrei particularmente em obter os dados;
- À Prof. Dra. **Maria de Rosário Martins** pelo suporte técnico e estatístico de que pude dispor e pelas valiosas sugestões que enriqueceram este trabalho e sobretudo pelo facto de que mesmo distante não mediu esforços para incentivar-me a seguir em frente, especialmente nos momentos em que me predispunha a mudar de tema;
- Aos meus pais, **Bhangy Cassy** e **Alexandra Rodrigues**, pelo exemplo de vida e por me terem mostrado desde cedo a importância da Educação;
- Aos meus irmãos, **Annette**, **Danilo** e **Bachir**, que vezes sem conta tiveram que partilhar momentos difíceis, próprios do percurso de estudante;
- À **Família Rodrigues**, pelo imensurável envolvimento ao longo da minha formação, que contribuiu de forma significativa para a minha progressão na carreira estudantil;
- A **Ayaz**, **Nuno Mamudo**, **Paulo Almeida**, **Alex**, **Mussagy**, **Yolanda Mambo**, **Cândido** de quem obtive particular apoio para a construção da base de dados usados neste trabalho;
- Aos meus amigos, **Agnalda**, **Sádia**, **Cintia** e **Cláudio** que comigo percorreram as etapas de crescimento desde a Escola à Universidade;
- Aos meus amigos do **Curso de Informática**, que sempre me motivaram a prosseguir, em especial à **Uneisa** e ao **Danilo Sousa**;
- Aos meus colegas do **curso de Estatística** que desde a minha entrada na Universidade em 2005 até 2008, contribuíram para que os 4 anos do Curso tivessem sido repletos de momentos agradáveis;
- À **dona Zulmira**, **Sr. Augusto** e ao **Eurico** que vezes sem conta tornaram mais fácil o dia-a-dia do **Curso de Estatística**;
- A todos os **docentes** do **DMI**;
- A todos os que contribuíram directa ou indirectamente para a realização deste estudo...

um bem-hajam!

Declaração de Honra

Declaro por minha honra que este trabalho é resultado da minha própria investigação, e que não foi submetido para outro grau que não seja o indicado- **Licenciatura em Estatística na Faculdade de Ciências da Universidade Eduardo Mondlane.**

Maputo, 18 de Fevereiro de 2009

Sheyla Ratan Rodrigues Cassy
(Sheyla Ratan Rodrigues Cassy)

Resumo

O risco de crédito constitui uma das principais preocupações das instituições financeiras que têm presenciado nos últimos anos um movimento de evolução nas técnicas de avaliação e gestão de risco de crédito com uma crescente na utilização de modelos quantitativos onde se destacam os modelos de *credit scoring*. Este trabalho teve como objectivo a criação de modelos de aprovação de clientes (*application scoring*) e outro, comportamental (*behavioural scoring*), pela aplicação da regressão logística, para avaliação do risco de crédito, numa instituição financeira. Esta instituição disponibilizou informação demográfica e sobre o comportamento, relativa a uma amostra dos seus clientes no período de Janeiro a Dezembro de 2003. Pelo método *forward stepwise* foi feita a escolha das variáveis que compuseram o modelo, gerando posteriormente o score para cada cliente, baseado nos coeficientes das variáveis escolhidas. Os modelos foram considerados eficientes na previsão da situação dos clientes, pois com uma percentagem de classificação correcta superior a 80% apresentaram uma óptima adequação em relação à discriminação entre bons e maus pagadores. Assim sendo, os modelos de *credit scoring* fornecem subsídios à instituição, auxiliando-a na prevenção e redução da inadimplência e consequentemente na diminuição dos custos operacionais.

Palavras chave: risco de crédito, modelos de *credit scoring*, regressão logística

ÍNDICE

CAPÍTULO - 1

1. <u>Introdução</u>	1
1.1 Contextualização	1
1.2 Definição do problema	3
1.3 Relevância do estudo	3
1.4 Objectivos do estudo	5
1.4.1 Objectivo Geral.....	5
1.4.2 Objectivos Específicos.....	5
1.5 Questão orientadora do estudo	5
1.6 Delimitações do estudo	6
1.7 Limitações do estudo	6
1.8 Considerações éticas	6
1.9 Estrutura do relatório de estudo	7

CAPÍTULO - 2

2. <u>Fundamentação teórica</u>	8
2.1 Conceito de Risco de Crédito	9
2.2 Análise de risco de crédito	9
2.3 Modelos de risco de crédito	11
2.3.1 Sistemas especialistas de análise subjectiva	12
2.3.2 Modelos de Credit Rating.....	12
2.3.3 Modelos de Credit Scoring	13
2.4 Técnica estatística	17
2.4.1 Regressão logística	18

CAPÍTULO - 3

3. <u>Metodologia</u>	24
3.1 Tipo de Estudo	24
3.2 Amostra	25
3.3 Definição das variáveis do estudo	27

3.4 Desenvolvimento e análise dos modelos	28
3.4.1 Estimação dos parâmetros dos modelos	28
3.4.2 Significância e Qualidade dos Modelos	29
3.5 Validação dos Modelos	32
CAPÍTULO - 4	
<u>4. Apresentação, análise e discussão dos modelos</u>	33
4.1 Caracterização da Amostra	33
4.2 Apresentação dos modelos	37
4.2.1 Modelo de Aprovação de crédito (<i>application scoring</i>).....	37
4.2.2 Modelo comportamental (<i>behavioural scoring</i>).....	42
4.3 Validação dos Modelos	46
4.3.1 Matriz de confusão ou tabela de classificação.....	46
4.3.2 Curva de ROC.....	48
CAPÍTULO - 5	
<u>5. Conclusões e recomendações para a análise de risco de crédito</u>	49
5.1 Conclusões	49
5.2 Possíveis contribuições para a análise de risco de crédito	50
<u>Bibliografia</u>	52
<u>Anexos</u>	56

LISTA DE GRÁFICOS

Gráfico 2.1: Função logística.....	20
Gráfico 4.1: Distribuição dos clientes que compuseram a amostra com base no género	33
Gráfico 4.2: Distribuição dos clientes que compuseram a amostra de acordo com o grupo etário ..	34
Gráfico 4.3: Curva de ROC do modelo de aprovação de crédito	48
Gráfico 4.4: Curva de ROC do modelo comportamental de crédito	48

LISTA DE FIGURAS

Figura 2.1: Representação esquemática do conceito de crédito	9
---	----------

LISTA DE TABELAS

Tabela 2.1: Interpretação da Curva de ROC.....	23
Tabela 4.1: Caracterização da amostra de acordo com o estado civil e o género.....	34
Tabela 4.2: Distribuição dos clientes que compuseram a amostra, por Sector Profissional.....	35
Tabela 4.3: Distribuição dos clientes que compuseram a amostra de acordo com o nível de formação académica	36
Tabela 4.4: Distribuição dos clientes que compuseram a amostra, por Agência.....	36
Tabela 4.5: Estatística descritiva relativa ao valor do crédito concedido	37
Tabela 4.6: Estimativas dos parâmetros do modelo de aprovação de crédito	37
Tabela 4.7: Estatística <i>GM</i> do modelo de aprovação de crédito.....	40
Tabela 4.8: Medidas <i>pseudo R²</i> do modelo de aprovação de crédito	40
Tabela 4.9: Tabela de contingência do teste Hosmer- Lemeshow do modelo de aprovação de crédito	41
Tabela 4.10: Teste de Hosmer-Lemeshow do modelo de aprovação de crédito	42
Tabela 4.11: Estimação dos parâmetros do modelo comportamental de crédito (<i>behavioural scoring</i>).....	42
Tabela 4.12: Estatística <i>GM</i> do modelo comportamental de crédito.....	44
Tabela 4.13: Medidas <i>pseudo R²</i> do modelo comportamental de crédito	45
Tabela 4.14: Tabela de contingência do teste Hosmer- Lemeshow do modelo comportamental de crédito	45
Tabela 4.15: Teste de Hosmer-Lemeshow do modelo comportamental de crédito.....	46
Tabela 4.16: Matriz de classificação do modelo de aprovação de crédito (% acertos).....	46
Tabela 4.17: Matriz de classificação do modelo comportamental de crédito (% de acertos).....	47
Tabela A-1: Variáveis consideradas na análise e sua natureza	56
Tabela A-2: Variáveis dummy.....	57

CAPÍTULO 1

1. INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Factores como o crescente custo de vida e uma maior disponibilidade de bens de consumo vêm levando a uma maior procura de concessão de crédito. Como em qualquer outro contexto, a concessão de crédito traduz-se na disponibilização de um valor presente mediante uma promessa de pagamento desse mesmo valor no futuro, que pressupõe a confiança na solvabilidade do devedor, ou seja, a confiança de que o mesmo irá cumprir com o acordado previamente. No entanto, nem sempre os clientes que obtêm o crédito honram com os seus compromissos e conseqüentemente, paralelamente ao aumento da solicitação de concessão de crédito, as instituições financeiras passaram a assumir uma exposição maior ao risco de inadimplência ou seja, o risco do não recebimento (parcial ou total) dos juros e prestações do crédito.

Assim, sendo a concessão de crédito um dos principais negócios presentes no quotidiano das instituições financeiras, é óbvia a necessidade de se proceder a uma análise cuidadosa das solicitações de concessão de crédito de forma a minimizar o risco em que as instituições incorrem quando tomam a decisão de conceder crédito. Essa avaliação do risco de crédito converteu-se ao longo dos tempos num processo organizado e cada vez mais sistematizado para analisar dados que possam de forma efectiva caracterizar o potencial cliente a quem se pode proceder à concessão de

crédito. Na prática, este processo visa sobretudo identificar os riscos para a instituição que concede o crédito, evidenciar conclusões quanto à capacidade de pagamento do solicitante e fazer recomendações sobre o melhor tipo de empréstimo a ser concedido.

É nesse contexto que uma adequada caracterização do solicitante é tida como uma contribuição não só para minimizar a possibilidade da instituição não reaver o dinheiro como também desta ter os seus lucros garantidos. No entanto, tendo em conta o universo dos solicitantes, a selecção dos clientes a quem se decide conceder o crédito requer o desenvolvimento de meios e instrumentos que permitam classificar de forma ágil e correcta o cliente. Por isso, a necessidade de controlo e gestão eficaz do risco tem levado a que as instituições financeiras passem a primar pelo aperfeiçoamento das técnicas utilizadas para essa função, e à sofisticação na avaliação do risco de crédito que é também reflexo da multiplicidade, qualidade e origem das informações disponíveis para o analista do crédito.

Cada vez com maior frequência se faz uso de técnicas estatísticas de classificação, bem como de modelos econométricos e matemáticos que têm aplicação em diversos fins, como seja a aprovação do crédito, a determinação do *rating* de crédito, a precificação de crédito e a estratégia de cobrança. Estes modelos fundamentam uma análise mais objectiva na avaliação de risco de crédito procurando particularmente minimizar o custo social provocado pelo incumprimento do acordado entre a instituição e o cliente.

Tal deve-se ao facto de que, tendo como base informações a partir de dados reais e em estatísticas, estes modelos apresentam-se mais seguros do que os métodos subjectivos ou de juízos de valor e habilitam as instituições a realizar milhares de avaliações de concessão de crédito de forma consistente e imparcial. Assim, ao longo dos tempos se vêm desenvolvendo modelos de *credit scoring* na procura de manter a qualidade do crédito concedido, independentemente do volume de operações a realizar e que são utilizados como ferramenta de suporte à tomada de decisão do gestor de crédito.

1.2 DEFINIÇÃO DO PROBLEMA

Tendo em conta a natureza da actividade das instituições financeiras, a maioria dos ambientes de avaliação de risco de crédito em Moçambique, exige que as decisões sejam tomadas rapidamente e de forma dinâmica. No entanto, a análise desse risco tem habitualmente como base uma enorme variedade de informações, vindas das mais diversas fontes acabando por constituir um volume de informação que pelas suas dimensões pode influenciar negativamente a capacidade de tomada de decisões. Para além de que, ao proceder-se à avaliação destas informações com base em critérios qualitativos, o carácter subjectivo torna-se presente e, muito embora até seja possível identificar os factores que influenciam as decisões, com frequência não se consegue explicar os processos de tomada das decisões.

Desta forma, é evidente que os múltiplos aspectos de cada informação e o número de informações a ser considerado na avaliação do risco presente na concessão de um crédito exigem que sejam criados mecanismos operacionais e métodos estatísticos, que de alguma forma, incorporem todos os dados relevantes num único número para que de forma mais objectiva se possa proceder à análise desse risco. Isto é, torna-se imperioso que o processo de tomada de decisão possa assentar em informação automatizada sustentada em instrumentos que de forma objectiva sirvam de orientação para a tomada de decisão.

1.3 RELEVÂNCIA DO ESTUDO

A probabilidade de inadimplência tem um papel importante na gestão do risco de crédito, auxiliando nas operações de crédito e no estabelecimento de limites de crédito. Por isso, a avaliação do risco de crédito torna-se num importante problema administrativo da área de análise financeira cuja solução passa por dispor de meios objectivos que possam habilitar as instituições a realizar milhares de avaliações de concessão de crédito de forma consistente e imparcial.

A crescente tendência de desenvolvimento de modelos de análise de risco de crédito é particularmente justificada pela necessidade de disponibilidade de mecanismos cada vez mais simples, eficazes e capazes de avaliar e conceder agilmente o crédito com pequena probabilidade de comprometer o nível de risco assumido pela instituição. Esses modelos permitem que a instituição financeira assuma o papel de fomentar e desenvolver económica e socialmente o País com um nível de risco que deve permitir a auto-sustentabilidade da actividade a que a instituição se dedica mostrando-se assim de vital importância, pois, qualquer classificação do cliente feita de forma inadequada pode causar prejuízos ou privar a instituição de lucros.

Embora estes modelos informatizados de análise deste risco sejam específicos para cada produto de crédito eles, além de fornecerem subsídios ao estabelecimento de parâmetros mínimos para concessão de crédito e gestão de riscos, possibilitam a definição de políticas de crédito diferenciadas em função das características e do porte do cliente que permitem a construção de uma base tanto para a correcta predição das operações, quanto para a definição de garantias adequadas a cada situação. Neste contexto, ao tomar-se em consideração o expressivo crescimento e a importância do mercado de créditos em Moçambique, a disponibilidade de modelos quantitativos para a tomada de decisão sobre a concessão de crédito como instrumento de apoio a qualquer que seja a metodologia actualmente adoptada na instituição a ser usada como local de estudo, contribuirá para facilitar a avaliação do risco do crédito.

Adicionalmente o processo de análise do risco de crédito que fundamenta a tomada de decisão em conceder ou não o crédito torna-se mais eficiente com mecanismos mais uniformes e credíveis que contribuam para reduzir a possibilidade de numa mesma instituição, a decisão de conceder ou não o crédito solicitado por um determinado cliente poder ser diferente; ser possível avaliar um maior número de solicitações por unidade de tempo e assim correr um menor risco dos clientes optarem por outra instituição (que procede de forma mais rápida) estimulando para estas, a retenção de clientes e o aumento de lucros e assim minimizar os riscos de inadimplência. Todos esses factores são acrescidos da escassez de pesquisas académicas no mercado nacional direccionadas à gestão do risco de crédito que permitem definir como oportuna a ocasião para o desenvolvimento deste estudo.

1.4 OBJECTIVOS DO ESTUDO

1.4.1 OBJECTIVO GERAL

O presente estudo tem como objectivo geral a criação de modelos de *credit scoring*, que possam através da análise de dados históricos, fazer uma previsão do comportamento futuro de clientes duma instituição financeira.

1.4.2 OBJECTIVOS ESPECÍFICOS

Com este estudo pretende-se:

- propor um modelo, através da aplicação da técnica de regressão logística, que permita com base nas informações cadastrais determinar o risco de crédito de cada cliente;
- identificar os factores lactentes que melhor discriminam os bons dos maus pagadores;
- propor um novo método de análise de risco de crédito para a instituição financeira.

1.5 QUESTÃO ORIENTADORA DO ESTUDO

Para responder aos objectivos, o estudo foi orientado para dar a resposta à seguinte questão:

No âmbito da concessão de crédito, quais os factores que permitem discriminar os bons dos maus pagadores?

1.6 DELIMITAÇÕES DO ESTUDO

Os dados usados no presente estudo foram obtidos numa base de clientes numa instituição financeira a operar no mercado moçambicano referente aos empréstimos concedidos entre Janeiro e Dezembro de 2003. A obtenção da base de dados só foi possível após o compromisso de que nem a instituição nem os clientes seriam identificados.

1.7 LIMITAÇÕES DO ESTUDO

Tendo em conta que é com base em informações passadas que são construídos os modelos a disponibilidade e a qualidade da base de dados são fundamentais para o sucesso do modelo. Assim a principal limitação do estudo está relacionada com a disponibilização numa base de dados fiável. O sistema de gestão de créditos usado pela instituição financeira agrega poucos dados sobre os clientes sendo que informações que são relevantes em modelos de *credit scoring* como por exemplo salário do cliente, número do agregado familiar, número de parcelas, não puderam ser consideradas para a construção do modelo pelo facto de não serem disponibilizadas pelo sistema computacional. De realçar também que neste estudo, foram vários os casos de clientes que nas variáveis profissão e educação tinham como opção “sem informação ou não indicada”. Estes casos não foram eliminados da amostra devido à sua elevada frequência, particularmente entre os maus pagadores.

1.8 CONSIDERAÇÕES ÉTICAS

Em nenhuma das partes deste trabalho é possível identificar a instituição financeira que concedeu a base de dados. A não identificação da instituição financeira da qual foram obtidas as informações é imposta pelos princípios éticos de relação comercial pelo que nenhum dado que pudesse identificar objectivamente o cliente estava contido na base disponibilizada garantindo-se, assim a confidencialidade da informação fornecida pelo cliente.

1.9 ESTRUTURA DO RELATÓRIO DE ESTUDO

O relatório do estudo está estruturado em 5 capítulos:

- **Capítulo 1:** Introdução - neste capítulo é apresentado o problema do estudo, os objetivos, a relevância e as limitações do mesmo;
- **Capítulo 2:** Fundamentação teórica - neste capítulo procura-se elaborar algumas considerações e definições sobre o risco de crédito e modelos de risco de crédito tendo como foco os modelos de *credit scoring*, as suas vantagens e desvantagens para as instituições que concedem crédito bem como a devida explicação do problema em causa;
- **Capítulo 3:** Metodologia onde apresenta-se a visão geral das técnicas adoptadas, neste caso, a regressão logística. A amostra do estudo é caracterizada, tal como a fonte de dados e as potenciais variáveis explicativas;
- **Capítulo 4:** Resultados e Discussão - procede-se à descrição da base de dados e à apresentação, análise e discussão dos modelos de *credit scoring* propostos;
- **Capítulo 5:** Conclusões e Recomendações – neste capítulo apresentam-se as principais conclusões e possíveis contribuições para o melhoramento da análise do risco de crédito na instituição.

2. FUNDAMENTAÇÃO TEÓRICA

A concessão de empréstimos é uma actividade comercial dos bancos que embora constitua um importante impulsionador da actividade económica, por disponibilizar recursos financeiros que permitem a pessoas físicas financiar as suas necessidades, não é habitualmente tomada em ambiente de total certeza em relação aos seus resultados. Tal como fundamenta Neto (2001), ao ser voltada para o futuro, a possibilidade de retorno do crédito não é totalmente previsível levando a que em todo o crédito ao consumidor esteja presente algum risco.

Se tal risco for mal avaliado, a instituição irá perder dinheiro, quer seja por aceitar clientes que irão gerar prejuízos, quer seja por recusar clientes bons que gerariam lucros (Gonçalves, 2005). Neste contexto, qualquer decisão menos adequada leva não apenas a prejuízos financeiros à instituição, mas também prejuízos financeiros e morais aos seus clientes, acabando assim por ter um impacto significativo nos lucros das instituições concessionárias de crédito.

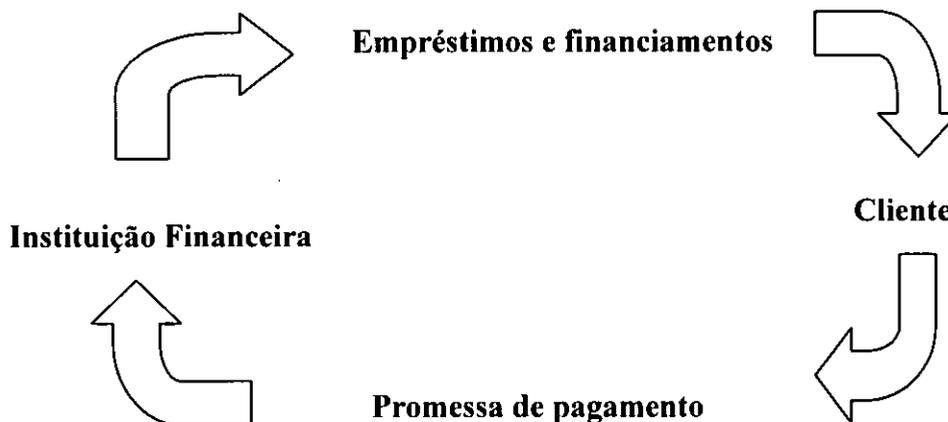


Figura 2.1: Representação esquemática do conceito de crédito

2.1. CONCEITO DE RISCO DE CRÉDITO

De acordo com Lima (2008), dentre as várias conceitualizações de risco de crédito, uma linha de raciocínio tem predominado e define o crédito como a troca de um valor presente por uma promessa de reembolso futuro, não necessariamente certa, em virtude do “factor risco”. Bernstein (1997) citado por Carvalho (2004), argumenta que risco (que deriva do italiano *riscare*, que significa ousar), é uma opção e não um destino e portanto pode ou não ser assumido. Assim, no caso das instituições financeiras, há necessidade de se proceder a avaliações da capacidade de quem solicita o empréstimo, de honrar com o compromisso com a instituição concessionária de forma a garantir que o risco em que a instituição incorre possa por esta ser assumido.

2.2 ANÁLISE DE RISCO DE CRÉDITO

A análise de risco de crédito, de acordo com Guimarães (2008), permite agrupar os tipos de risco em duas formas principais (i) o risco que o solicitante de crédito apresenta e (ii) o risco que a

instituição credora tem condições para aceitar. Se o risco inerente à instituição depende da política adoptada por esta, já o risco apresentado pelo solicitante é definido pelas suas características e por isso se apresenta como de extrema importância no processo de concessão de crédito.

Para que o risco de crédito possa ser avaliado é necessário que sejam considerados factores que permitam decidir se o solicitante será ou não capaz de honrar os seus compromissos, pelo que quando um dado cliente procura obter crédito, as instituições concessionárias de crédito, regra geral, solicitam o preenchimento dum ficha cadastro. Este processo de análise em que se procura avaliar não só a capacidade financeira mas, igualmente, o perfil de risco apresentado pelo solicitante, pode segundo Lima (2008), fazer uso de técnicas de cariz subjectivo, (que assenta num juízo de valor e por isso dependente da experiência adquirida, disponibilidade de informações e sensibilidade de cada analista quanto à aprovação do crédito) e técnicas de cariz objectivo também designadas por técnicas estatísticas por basear-se em processos estatísticos.

Mesmo tendo em conta que a análise subjectiva do risco de crédito seja dependente da sensibilidade de cada analista, é através da experiência do agente de crédito que se torna possível identificar factores relacionados com o carácter, capacidade, capital e condições de pagamento. Por isso cada vez mais se usam modelos de análise em que se associam as técnicas subjectivas às técnicas objectivas para que as regras de análise sejam mais consistentes, que o processo de decisão possa ser descentralizado e o tempo precioso dos profissionais de crédito fique reservado para um grupo de processos habitualmente considerado marginal, onde o seu *know-how* será melhor rentabilizado, enquanto que os modelos estatísticos encarregar-se-ão das aprovações de transacções de baixo risco, definindo condições e limites de crédito diferenciados e previamente estabelecidos (Araújo, 2006).

2.3 MODELOS DE RISCO DE CRÉDITO

Os modelos de risco de crédito são modelos que servem para quantificar a possibilidade da instituição concessionária incorrer em perdas. Segundo Araújo (2006), algoritmos, fórmulas, sistemas ou regras que procurem representar processos ou atributos reais relacionados com o risco de crédito podem ser considerados modelos de risco de crédito que devido às suas características procuram facilitar a compreensão dum determinado fenómeno e eventualmente proceder à sua exploração.

Na extensa literatura sobre os modelos de risco de crédito são apresentadas várias formas de subdivisão dos modelos com base no facto de que ao longo dos tempos as instituições bancárias que concedem créditos terem desenvolvido diferentes modelos para se proceder à análise de risco de crédito. No presente trabalho adoptou-se a classificação de Saunders (2000) citada por Araújo (2006) que divide os modelos em: (i) sistemas especialistas de análise subjectiva, (ii) modelos de *credit scoring* e (iii) modelos de *credit rating*.

Tal como sugeriu Pereira (2004), outros modelos para análise de risco de crédito incluem modelos para estimar a probabilidade do cliente pagar o empréstimo que já está em atraso (*collection scoring*), defraudar a instituição (*fraud scoring*), comprar um produto após uma campanha de marketing (*propensity scoring*), cancelar a conta ou um produto (*attrition scoring*), os modelos que têm como objectivo ordenar os clientes quanto à probabilidade de ter um problema de crédito em pelo menos um produto, dentro dum prazo pré determinado (*customer scoring*) e os modelos *profit scoring* que objectivam ordenar os clientes quanto à probabilidade de dar lucro à instituição, que têm em comum as metodologias de análise usadas nos modelos de aprovação de crédito (*application scoring*) e comportamental de crédito (*behavioural scoring*).

2.3.1 SISTEMAS ESPECIALISTAS DE ANÁLISE SUBJECTIVA

Tal como a própria designação indica, os sistemas especialistas de análise subjectiva são baseados num julgamento subjectivo realizado pelo analista de crédito, isto é, esta análise assenta basicamente na experiência do analista. Para a concretização desta análise o profissional atribui pontuações aos factores por si identificados como relevantes, dentro do conjunto de informações que o cliente faculta quando preenche a ficha cadastral ou mesmo por entrevista, tendo como base a sua experiência e a própria percepção sobre todo o processo de concessão de crédito.

Segundo Carvalho (2004), comumente o analista procede à análise dos chamados C's de créditos que incluem carácter, capacidade, capital e condições de pagamento. Esta forma de análise é considerada como um processo muito demorado pelo que não permite a análise de muitos pedidos. Por outro lado, há que ter em conta o facto de diferentes analistas poderem apresentar diferentes decisões relativas a um mesmo cliente reduzindo a viabilidade da tomada de decisão pela falta de uniformidade e tornando a instituição mais dependente da percepção de risco e experiência profissional dos analistas.

2.3.2 MODELOS DE CREDIT RATING

Os modelos de *credit rating* são também designados por sistemas de classificação, pois são usados para classificar os empréstimos em categorias de risco de crédito tendo como base a probabilidade de perda, definida pela atribuição de pontuações que permitem que se proceda ao agrupamento dos créditos (Araújo, 2006).

2.3.3 MODELOS DE CREDIT SCORING

Por *credit scoring* entende-se um sistema ou ferramenta utilizado por instituições que concedem crédito, para verificar se o crédito deve ou não ser concedido ao cliente ou para proceder ao acompanhamento e administração dos créditos já concedidos, gerando um banco de dados para as instituições. A principal característica deste tipo de modelo é reduzir o tempo de processamento da decisão de crédito, disponibilizando o gerente de crédito para outras tarefas, como por exemplo: avaliação de outras características não computadas pelo sistema; captação de novos clientes; e recuperação de créditos problemáticos já concedidos.

Os modelos de *credit scoring* procuram definir a probabilidade dum cliente vir a ser bom pagador ou mau pagador com base em informações contidas na base de dados referentes aos clientes existentes. Para tal atribuem-se pontuações às variáveis de decisão de crédito dum proponente, mediante a aplicação de técnicas estatísticas. Estes modelos de *credit scoring* identificam os factores-chave que influenciam na adimplência e inadimplência dos clientes e têm por base uma medida (score) absoluta (pontuação) ou relativa (percentil), que indica a probabilidade de, num determinado período de tempo (geralmente um três anos), um cliente incorrer num comportamento não desejado (como atraso ou incumprimento).

Sem dúvida que com um modelo de *credit scoring* é possível classificar os solicitantes quanto ao grau de risco de crédito; caracterizar a relação do cliente com a instituição financeira; adoptar políticas de crédito concorrenciais (dado o automatismo que assegura) para além de estimular a uma maior interactividade de vários departamentos da instituição na definição da política de crédito, muito embora os modelos de *credit scoring* sejam específicos para a aprovação de cada produto de crédito, (Gonçalves, 2005).

Segundo Régis (2007), os modelos de *credit scoring* podem ser divididos em modelos de aprovação de crédito (*application scoring*) e modelos comportamentais de crédito (*behavioural scoring*). Os modelos de aprovação de crédito (*application scoring*) são utilizados para auxiliar instituições

financeiras na tomada de decisão de concessão de crédito a um novo cliente fazendo uso de informações constantes do cadastro do cliente e de acordo com Pereira (2004), o seu principal objectivo é estimar a probabilidade de um indivíduo que está a solicitar o crédito se tornar inadimplente (mau pagador) antes de completar um período pré-fixado, isto é, não cumprir com a sua promessa de pagamento no período acordado.

Num estudo realizado por Araújo (2006) o modelo de aprovação assentou nas variáveis referentes a número de parcelas, valor do empréstimo, tempo de funcionamento e o resultado líquido do negócio enquanto que o modelo comportamental usou as variáveis número de parcelas, valor do empréstimo, tempo de funcionamento, resultado líquido do negócio, número de dependentes, não casados e o avalista não possui crédito anterior. Similarmente para Gonçalves (2005) as variáveis que se mostraram mais relevantes na análise de risco de crédito foram o género, estado civil, tempo no actual emprego, número de parcelas, valor total do empréstimo, tempo na residência actual, tipo de crédito, profissão, percentagem do valor do empréstimo sobre o salário e primeira aquisição embora alguns factores tenham efeito positivo e outros negativos sobre a pontuação ou score de crédito.

O modelo comportamental (*behavioural scoring*) incorpora variáveis que retratam a história do cliente com a instituição. Esse modelo desenvolvido para estimar a probabilidade de um cliente que já possui um determinado produto ter problema de crédito nos n meses seguintes, objectiva auxiliar o analista de crédito nas suas decisões sobre renovações de empréstimos de clientes, renegociações de dívidas, determinação do montante para renovação de empréstimo, determinação do valor da parcela na renovação de empréstimo, enfim, todas as decisões relativas à gestão do crédito de clientes que já possuem uma relação ou seja um histórico com a instituição.

Por outro lado, o modelo comportamental é tido como com mais vantagens do que o modelo de aplicação pelo facto de possuir um número maior de variáveis disponíveis para o seu ajuste. Além das variáveis disponíveis no momento da concessão já se conhece o comportamento de uso do produto pelo cliente. Dessa forma é possível obter modelos com poder de discriminação bem

superior aos observados em modelos de aplicação com a utilização geral, das mesmas metodologias.

2.3.3.1 VANTAGENS E DESVANTAGENS DOS MODELOS DE *CREDIT SCORING*

Embora diante da maior agilidade na decisão, menor custo, maior objectividade e até mesmo melhor poder preditivo, os modelos de *credit scoring* tenham aos poucos se popularizado e tornando-se actualmente largamente aplicados (Guimarães, 2008), independentemente da ferramenta estatística usada para o seu desenvolvimento, são descritas na literatura vantagens e desvantagens que estão presentes na elaboração desses modelos. Algumas dessas vantagens e limitações/desvantagens enumeradas por Chorão (2005) e Pereira (2004) são as seguintes:

Vantagens

- Consistência na análise das solicitações de créditos: modelos bem elaborados, que utilizam a experiência e habilidade da instituição a vários níveis, servem para administrar objectiva e justamente os créditos de clientes e candidatos a estes, devido à condição da sua revisão;
- Melhor organização da informação de crédito: em conjunto com a consistência, a melhoria organizacional é fundamental para que a informação de crédito seja uma consequência natural do processo, permitindo o aumento do lucro das instituições e economia de tempo e mão-de-obra;
- Uso mais eficiente de fontes de dados de terceiros: assegura o uso de informações que habitualmente eram somente arquivadas;

- Eliminação de aproximações subjectivas em excesso: devido a necessidades legais, a influência do subjectivismo é minimizada;
- Melhor entendimento do processo: a construção do modelo é uma experiência de aprendizagem da política de crédito da instituição permitindo a identificação de variáveis tidas como mais importantes na discriminação dos clientes;
- Melhoria da performance: foca os esforços dos gestores de crédito, significando mais eficiência e melhor uso do tempo favorecendo a agilidade adquirida pela instituição financeira que concede o crédito.

Limitações/Desvantagens

- Custo de desenvolvimento: desenvolver um sistema de *credit scoring* acarreta custos não só com o sistema em si, mas com o suporte necessário, como equipamentos, consultoria e outros;
- Excesso de confiança nos modelos: utentes inexperientes podem ver o modelo como “sem defeitos e preciso” e não avaliar de forma crítica os seus resultados;
- Falta de qualidade ou de dados oportunos: se o modelo necessita de dados que não são comuns aos clientes ou não foram informados, pode causar problemas na sua utilização e gerar resultados diferentes dos esperados ou, até mesmo, a rejeição destes;
- Dificuldade no entendimento do significado dos scores: a falta de treinamento e aprendizagem de como usar os dados do modelo podem ocasionar problemas sérios para a instituição, devido ao uso inadequado do sistema;

- Permanente actualização: os modelos devem ser actualizados constantemente de modo a não degradarem o poder preditivo destes.

2.4 TÉCNICA ESTATÍSTICA

Segundo Neto e Carmona (2007), as três principais técnicas para elaboração de modelos de *credit scoring* são a análise discriminante, a regressão logística e, mais recentemente modelos baseados em redes neurais. Quando comparada com as técnicas conhecidas em regressão, a regressão logística distingue-se essencialmente pelo facto da variável explicativa ser categórica. Como método de predição para variáveis categóricas, a regressão logística é comparável às técnicas supervisionadas propostas na aprendizagem automática (árvores de decisão, redes neuronais, etc.), ou ainda à análise discriminante preditiva em estatística exploratória. A regressão logística pode ser usada quando a variável dependente é não-métrica e as variáveis independentes são métricas ou não-métricas. Nesta situação o pesquisador está interessado na previsão e explicação do impacto que as variáveis independentes geram na variável dependente. Mais especificamente, o pesquisador procura saber como as categorias da variável dependente estão associadas às variações das variáveis independentes.

Hair *et al.* (2005) apontam uma lista de motivos que levam a optar pela regressão logística mesmo quando a variável possui apenas duas categorias. Primeiro, a regressão logística é menos afectada quando as matrizes de covariância (para os grupos da variável categórica) não são iguais o que constitui uma premissa básica da análise discriminante. Segundo, a regressão logística pode lidar com variáveis independentes categóricas facilmente, enquanto que na análise discriminante o uso de variáveis *dummy* cria problemas com a variância ou covariância entre os grupos determinados pelas categorias da variável dependente. Finalmente, os resultados da regressão logística são similares aos da regressão múltipla nos termos de sua interpretação e medida dos diagnósticos válidos caso a caso para exame dos resíduos.

Outro aspecto que favorece a utilização da regressão logística é que os seus resultados podem ser interpretados em termos de probabilidade. Esse factor mostra-se particularmente importante nos modelos de risco de crédito, pois possibilita que seja medida a probabilidade de um determinado tomador assumir a condição de solvente ou insolvente, face um conjunto de atributos (Penha, 2002). Tendo em conta todos esses factores, a regressão logística foi a técnica escolhida para a discriminação e classificação dos grupos (bons e maus clientes ou adimplentes e inadimplentes), para este estudo.

2.4.1 REGRESSÃO LOGÍSTICA

2.4.1.1 DESCRIÇÃO DO MODELO

Neto e Brito (s/d) definem regressão logística, ou análise *logit*, como uma técnica de análise multivariada, apropriada para as situações nas quais a variável dependente é categórica e assume um entre dois resultados possíveis (binária). O objectivo desta técnica é gerar uma função matemática, cuja resposta permita estabelecer a probabilidade de uma observação pertencer a um grupo previamente determinado, em razão do comportamento de um conjunto de variáveis independentes. Como técnica estatística, a regressão logística é uma técnica usada para estudar a relação entre uma variável categorizada de interesse e um conjunto de outras disponíveis no estudo. Uma primeira qualidade da regressão logística está no facto da ligação *logit* produzir valores ajustados que variam entre 0 e 1, propriedade bastante importante no estudo de dados binários.

Na técnica de regressão logística, as previsões de pertinência ou classificação dos indivíduos em cada um dos grupos em estudo, são realizadas através da previsão directa da probabilidade do evento acontecer. No caso da variável dependente Y assumir apenas dois possíveis estados (0 ou 1) e haver um conjunto de n variáveis independentes X_1, X_2, \dots, X_n , o modelo de regressão logística pode ser escrito da seguinte forma:

$$P_i = \Pr [Y_i=1|X=X_i] = \frac{1}{1 + e^{-Z_i}}$$

Onde:

$$Z_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}$$

sendo:

- P_i = probabilidade de um evento ocorrer;
- Z_i = score (variável dependente);
- β_0 = constante;
- $\beta_1, \beta_2, \dots, \beta_n$ = coeficientes expressos em logaritmos;
- X_i = variáveis independentes.

A regressão logística calcula directamente a probabilidade condicional da variável dependente ser igual a 1. Devido à natureza não linear dessa transformação, a estimação dos coeficientes $\beta_0, \beta_1, \dots, \beta_n$ é feita através do Método da Máxima Verosimilhança. Para definir resultados delimitados por 0 e 1, a regressão logística usa uma relação entre as variáveis independentes e a variável dependente que se assemelha a uma curva em forma de S.

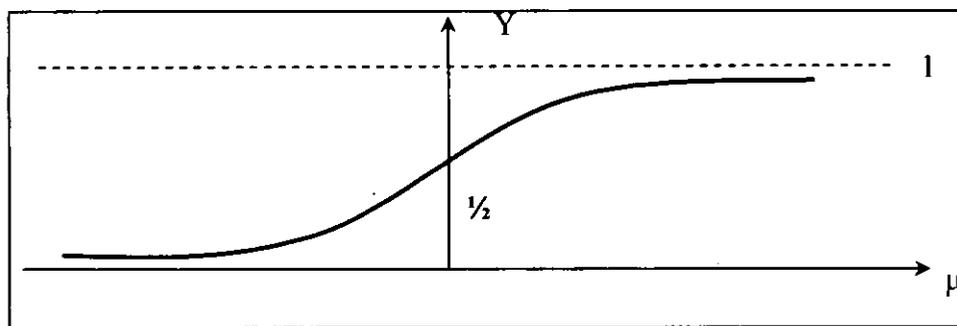


Gráfico 2.1: Função logística

Em níveis muito baixos da variável independente, a probabilidade tende a zero. À medida que a variável independente aumenta, a probabilidade aumenta para cima da curva, mas em seguida a inclinação começa a diminuir, de forma que em qualquer nível da variável independente, a probabilidade irá tender a 1, sendo que jamais irá exceder esse valor. A partir do valor dicotômico, o procedimento prevê a sua estimativa de probabilidade de que o evento ocorrerá ou não.

2.4.1.2 ESTIMAÇÃO DO MODELO

No método de regressão logística (e em outros métodos de regressão), a selecção das variáveis a reter pode ser feita de forma automática, de acordo com determinadas regras e algoritmos de selecção. Os métodos de busca sequencial procuram estimar a equação de regressão com um conjunto de variáveis e então seleccionar variáveis para acrescentar ou eliminar do modelo até que alguma medida de critério geral seja alcançada. Este tem por objectivo seleccionar variáveis que maximizam a previsão com o menor número de variáveis empregadas.

Existem dois tipos de abordagem de busca sequencial: (1) estimação *stepwise* e (2) adição *forward* e eliminação *backward*. A estimação *stepwise* (por etapas) permite examinar a contribuição de cada variável independente para o modelo. Estas variáveis são então seleccionadas para inclusão com base na sua contribuição incremental sobre as variáveis já presentes na equação. (Hair *et al.*, 2005). Uma grande distinção entre o método *stepwise* e os procedimentos de adição *forward* e eliminação

backward está no facto de que o primeiro método tem grande habilidade em acrescentar e eliminar variáveis em cada estágio, enquanto que nos dois últimos, depois de adicionada ou retirada uma variável da equação o modelo não tem como reverter este quadro.

Assim, tal como fundamenta Marroco (2007), no método *forward stepwise* inicia-se o modelo sem variáveis preditivas sendo as variáveis acrescentadas no modelo, uma a uma e usando critério de selecção previamente definido a variável acrescentada é mantida ou não no modelo, consoante verifique-se ou não o critério. No método *backward stepwise*, inicia-se com o modelo completo, isto é, com todas as variáveis da base de dados, e estimam-se sucessivos modelos em que as variáveis independentes são retiradas, uma a uma com base em determinado critério de exclusão.

O método *stepwise* é baseado num algoritmo estatístico que avalia a importância de cada variável independente. A importância de cada variável é definida em termos de medida de significância estatística do seu coeficiente. Este procedimento inicia através da estimação de um modelo apenas com o intercepto, inclui uma a uma, as variáveis mais significativas no modelo, e exclui aquelas que na presença de outras deixam de ter importância e termina quando nenhuma variável puder ser incluída ou excluída no modelo de acordo com os níveis de significância pré-estabelecidos. Os parâmetros utilizados habitualmente são 5% de significância para a entrada das variáveis e 10% para a saída.

Na regressão logística, é utilizada a estatística de Wald para testar a significância da equação estimada (Hair *et al.*, 2005). Ela fornece a significância estatística de cada coeficiente estimado, de modo que o teste de hipóteses pode ocorrer como acontece na regressão múltipla e assim permite verificar o impacto de cada variável independente na dependente ou seja testar a significância dos coeficientes do modelo.

2.4.1.3 AVALIAÇÃO DO AJUSTE DO MODELO E DA SUA CAPACIDADE PREDITIVA

Uma vez definido o modelo será necessário testar a sua qualidade. A avaliação do nível de ajuste do modelo é feita por meio do valor de *Likelihood* ($-2LL$), dos *pseudo* R^2 e do teste de Hosmer-Lemeshow. De acordo com Hair *et al.* (2005), um modelo bem ajustado terá um valor pequeno para $-2LL$, sendo o seu valor mínimo igual a zero (0), enquanto que para as medidas de adequação de ajuste de Cox-Snell R^2 e Nagelkerke R^2 , maiores valores indicam melhor ajuste do modelo. O Cox-Snell R^2 baseia-se no valor de *Likelihood* e tem uma escala que começa em zero (0) mas não alcança um (1) no seu limite superior. O Nagelkerke R^2 é similar ao Cox-Snell R^2 , porém tem uma escala que vai de zero (0) a um (1).

O teste de Hosmer-Lemeshow avalia as diferenças entre as classificações previstas pelo modelo e as observadas. Se as diferenças forem significativas, o grau de acurácia do modelo não é bom. O teste divide os casos em classes e compara as frequências previstas e observadas em cada classe por meio de uma estatística Qui-quadrado (Marroco, 2007).

Após a construção do modelo, é também necessário avaliar a capacidade preditiva do modelo, o que é feito através da matriz de classificação de forma a identificar a quantidade de clientes classificados correcta ou incorrectamente pelos modelos. Um segundo procedimento utilizado para avaliar a performance do modelo é a construção duma Curva ROC (*Receiver Operating Characteristic*).

Tal como ressalta Marroco (2007), a Curva ROC constitui uma técnica bastante útil para validar modelos de risco de crédito e está baseada nos conceitos da sensibilidade e da especificidade. Por isso, para a construção da Curva ROC, são calculadas a sensibilidade (proporção de acerto na previsão da ocorrência de um evento nos casos em que ele de facto ocorreu, que neste estudo indica a proporção de inadimplentes classificados como adimplentes) e a especificidade (proporção de acerto na previsão da não ocorrência de um evento nos casos em que ele de facto não ocorreu no caso, proporção de adimplentes bem classificados), para todas as observações da amostra, considerando diferentes pontos de corte do modelo.

A partir da determinação da especificidade e da sensibilidade esperada em cada valor previsto do modelo, é possível construir um gráfico em que a sensibilidade da previsão é observado no eixo das

abscissas e o índice da especificidade da previsão é observado no eixo das ordenadas, para cada um dos valores previstos possíveis do modelo.

A pontuação do modelo *logit* varia entre]0,1[podendo um determinado indivíduo ser pertença de uma ou outra população dependendo do *cutoff* considerado para avaliação. Sem dúvida que quanto maior a probabilidade extraída do modelo *logit* maior a probabilidade do indivíduo passar a mau pagador, pelo que à medida que se selecciona um *cutoff* mais baixo, maior a sensibilidade (aumentando-se a probabilidade dum indivíduo, através do modelo, ser classificado como inadimplente quando o é efectivamente) e menor a especificidade (diminuindo a probabilidade do indivíduo ser adimplente, por aplicação do modelo, sendo-o realmente).

A área sob a curva mede a capacidade de discriminação do modelo. De acordo com Marroco (2007) quanto melhor o poder de discriminação do modelo, mais próxima de 1 será a área abaixo da curva ROC podendo os valores para avaliação do modelo ser interpretados como:

Área ROC	Poder discriminante do modelo
0.5	Sem poder discriminante
]0.5;0.7[Discriminação fraca
]0.7;0.8[Discriminação aceitável
]0.8;0.9[Discriminação boa
≥ 0.9	Discriminação excepcional

Tabela 2.1: Interpretação da Curva de ROC

3 METODOLOGIA

3.1 TIPO DE ESTUDO

A classificação do tipo de estudo toma classicamente em consideração a abordagem qualitativa ou quantitativa do problema, os objectivos de pesquisa e os procedimentos técnicos adoptados. Usando a classificação sugerida por autores como Gil (1995), sob o ponto de vista de abordagem do problema este estudo é considerado quantitativo (procede a classificações e análises a partir de dados numéricos ou transformados em números); explicativo, quanto aos objectivos do estudo (procura identificar os factores que explicam a inadimplência integrando-os num modelo quantitativo) e estudo de caso ao tomar-se em consideração as técnicas e estratégias usadas (aborda de forma aprofundada, aspectos da gestão do risco de crédito no banco onde foi realizado o estudo).

3.2 AMOSTRA

Para o cálculo do tamanho da amostra, foi usada a fórmula que permite a determinação do tamanho estatisticamente significativo para amostra proveniente de população finita, quando a principal variável em consideração é nominal ou ordinal que se exprime como:

$$n = \frac{Z^2 \times p \times q \times N}{d^2(N-1) + Z^2 \times \hat{p} \times \hat{q}}$$

Onde:

N = tamanho da população que para o estudo foi considerado como a população de clientes no ano de 2003 que era de 10899;

Z = abscissa da normal padrão cujo valor é de 1.96 correspondendo a nível de significância de 95% que se mostrou possível de ser empregue na população em estudo;

d = erro amostral de 0.04 (4%), que representa a máxima diferença que se pode suportar entre a proporção amostral e a proporção populacional de maus clientes (inadimplentes), ou seja,

$|p - \hat{p}| \leq d$ tendo em conta que a maioria das pesquisas de determinação do tamanho da amostra apresenta uma margem entre 4% e 10%;

\hat{p} = estimativa da proporção de clientes inadimplentes na amostra que foi considerada como 0,5 que é o valor de proporção que fornece o maior tamanho da amostra, mantendo os outros elementos constantes;

\hat{q} = estimativa da proporção de clientes adimplentes 0,5 obtida por $(1 - \hat{p})$.

Desta forma, fazendo uso dos valores supra explicados e empregando a fórmula para o cálculo do tamanho da amostra, obteve-se como tamanho mínimo da amostra para que fosse representativa da população, um total de 592 clientes. Com a base e a variável resposta definida foram seleccionados aleatoriamente, a partir do universo de clientes da instituição financeira em estudo, 500 contratos de créditos tidos como bons e 500 considerados maus, realizados no período de Janeiro a Dezembro de 2003.

Embora no universo da base de dados fornecida pela instituição usada para o estudo, a proporção de bons clientes versus maus clientes tenha sido respectivamente 94.57% e 5.43%, neste trabalho, optou-se pela alternativa de uma amostra em proporções iguais (50%-50%) para que fossem propiciados resultados robustos, ao garantir que o modelo apreendesse as características dos dois subconjuntos presentes na amostra e se evitasse qualquer possível viés devido à diferença de tamanhos.

Este procedimento foi posterior à análise dos créditos constantes na amostra com o intuito de verificar a condição de adimplência nos 3 anos de vigência do crédito. Uma alternativa a esta técnica de composição da base de dados para o estudo seria extrair uma amostra aleatória do universo e depois ponderar os pesos de bons e maus clientes de acordo com a sua proporção na amostra, técnica também usada por diferentes autores como por exemplo Gonçalves (2005) em trabalhos similares.

3.3 DEFINIÇÃO DAS VARIÁVEIS DO ESTUDO

A identificação inicial das variáveis explicativas foi orientada por estudos anteriores sobre o assunto e resultou na identificação de variáveis como género, idade, estado civil, profissão, salário do solicitante, tempo de serviço, experiência profissional, valor do empréstimo, número de parcelas, valor das parcelas, número de atrasos, duração do maior atraso, data de desembolso (Vide Anexo, Tabela A-1).

É importante ressaltar que com o uso da técnica estatística de regressão logística foi necessário proceder-se a algumas manipulações em determinadas variáveis com o intuito de viabilizar o posterior processamento e aplicação da técnica estatística.

Tendo em conta que algumas das variáveis são de natureza qualitativa, como é o caso do género, idade, nível académico, etc., e o facto das mesmas influenciarem directamente a variável dependente, foi necessário o uso de variáveis *dummy* para inserir informações dessa natureza como variáveis explicativas nos modelos.

Segundo Gujarati (2000), as variáveis *dummy* ou categóricas consistem em variáveis construídas artificialmente para medir a presença ou não de algum atributo de natureza qualitativa. Elas assumem valor 1 ou 0, dependendo, respectivamente da presença ou não de determinado atributo.

De acordo com o autor supra citado, ao trabalhar-se com variáveis categorizadas deve-se seguir a regra de que o número de dummies para a variável seja sempre um a menos que o número de categorias da variável. Assim, foram inseridas N-1 dummies para representar as N categorias de cada variável qualitativa (Vide Anexo, Tabela A-2).

Para a construção do modelo de aprovação de crédito neste estudo usaram-se as variáveis: género, idade, estado civil, nível académico e sector profissional. E para a construção do modelo

comportamental de crédito foram adicionadas as variáveis: número de créditos anteriores e o histórico com a instituição.

3.4 DESENVOLVIMENTO E ANÁLISE DOS MODELOS

3.4.1 ESTIMAÇÃO DOS PARÂMETROS DOS MODELOS

Na técnica de regressão logística, as previsões de pertinência ou classificação dos indivíduos em cada um dos grupos (adimplentes ou inadimplentes), são realizadas através da previsão directa da probabilidade do evento de inadimplência acontecer. A variável dependente representa a “situação do cliente *i*”, a regressão logística calcula directamente a probabilidade condicional desta variável ser igual a 1, que, neste estudo, significa a probabilidade do cliente ser inadimplente. O ponto de corte adoptado foi 0.5, valor padronizado para a técnica de regressão logística. Esse valor de 0.5, conforme ressalta Hair *et al.* (2005), representa a probabilidade de ocorrência do evento segundo o critério de aleatoriedade ou chances iguais. Assim, aqueles clientes para os quais a probabilidade estimada de inadimplência resultou em valor inferior a 0.5, foram classificados como bons clientes. E aqueles para os quais a probabilidade de inadimplência foi superior a 0.5, foram classificados como maus clientes.

Procedeu-se à selecção das variáveis explicativas para compor o modelo usando o método *forward stepwise* pelo critério do menor *Likelihood Ratio*. A importância de cada variável foi definida em termos de medida de significância estatística do seu coeficiente. As variáveis que compõem os modelos fornecem informação acerca da contribuição ou importância do valor preditivo de cada variável. Foi considerado o sinal (positivo ou negativo) do valor dos coeficientes pois estes indicam a direcção da influência da relação (que factores aumentam a predisposição para a inadimplência e

que factores a diminuem). Assim os valores negativos dos coeficientes indicam que um aumento na variável independente resultará numa diminuição da probabilidade do caso se tornar inadimplente.

Por fim, atendendo à comparação que se pretende efectuar na identificação dos atributos que deverão constar num determinado modelo de *credit scoring*, calcula-se o rácio entre os *odds-ratio* (Exp (B)). A semelhança entre a noção de probabilidade e de *odds*, aplica-se igualmente, nesta definição. OR é, pois, a razão entre o *odd* de um atributo comparado com o *odd* de outro atributo, traduzindo quão forte está uma variável associada com a variável de interesse, quando comparada com as demais variáveis (Chorão e Martins, s/d). Assim sendo, quando o *odd* é maior que um (1), então o acontecimento, no presente caso a inadimplência, é mais provável acontecer; quando menor que um (1), é menos provável que aconteça.

3.4.2 SIGNIFICÂNCIA E QUALIDADE DOS MODELOS

3.4.2.1 TESTES DE SIGNIFICÂNCIA GLOBAL DO MODELO

Teste GM

A estatística *GM* segue uma distribuição do χ^2 com $(p-1)$ graus de liberdade, em que p é o número de parâmetros do modelo. Para a aplicação deste teste consideraram-se as seguintes hipóteses:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_{p-1} = 0$$

$$H_a: \text{existe pelo menos } 1\beta_i, i=0, \dots, p-1 \text{ diferente de zero}$$

Logo a regra de decisão define que

Se o valor $p \leq \alpha$ rejeita-se H_0 , concluindo-se que pelo menos uma das variáveis independentes do modelo consegue prever o *logit*.

Teste Hosmer-Lemeshow

No teste de Hosmer-Lemeshow, que é definido pela comparação das classificações previstas com as classificações observadas pelo modelo, se tais diferenças forem significativas, pode-se concluir que o grau de acurácia do modelo não é bom. Para este teste assume-se como hipóteses:

H_0 : O modelo ajusta-se aos dados;

H_a : O modelo não se ajusta aos dados.

A regra de decisão deve ser tida como:

Se o valor $p \leq \alpha$ rejeita-se H_0 , concluindo-se que o modelo não se ajusta aos dados.

Teste de significância dos coeficientes do modelo

Para testar a significância dos coeficientes do modelo, utiliza-se a estatística de Wald. A estatística Wald tem distribuição χ^2 , sendo calculada pelo quadrado da razão entre o coeficiente e o seu erro padrão. Os resultados deste teste para o modelo proposto indicam que os coeficientes das variáveis independentes seleccionadas são estatisticamente diferentes de zero. Assumindo-se como hipóteses:

H_0 : $\beta_i = 0, \beta_0, \beta_1, \dots, \beta_{p-1}$

H_a : $\beta_i \neq 0, \beta_0, \beta_1, \dots, \beta_{p-1}$

A regra de decisão será:

Se o valor $p \leq \alpha$ rejeita-se H_0 , concluindo-se que os coeficientes são significativos.

Medidas de associação múltipla entre as variáveis independentes e a variável dependente: Pseudo R^2

Os valores de R^2 Cox-Snell e R^2 Nagelkerke indicam em que extensão a variação da variável dependente (situação do cliente) é explicada pelo modelo (de um mínimo de 0 a aproximadamente um máximo de 1). Assim sendo, um modelo bem ajustado terá um valor pequeno para $-2LL$, sendo o seu valor mínimo igual a zero. Maiores valores de R^2 Cox-Snell (R^2 CS) e de R^2 Nagelkerke (R^2 N) indicam melhor ajuste do modelo.

O R^2 para o modelo *logit* pode ser calculado como:

$$R^2 \text{ logit} = \frac{(-2LL_{\text{nulo}} - (-2LL_{\text{modelo}}))}{-2LL_{\text{nulo}}}$$

O R^2 Cox-Snell é baseado no *log-likelihood* nunca atinge o valor máximo de um (1), considerando-se uma boa aderência valores acima de 0.22. Nagelkerke propôs um ajustamento ao R^2 Cox-Snell de forma a que o valor 1 possa ser atingido. Valores de R^2 Nagelkerke acima de 0.30 são considerados tradutores de boa qualidade de ajustamento do modelo.

3.5 VALIDAÇÃO DOS MODELOS

Para a validação do modelo de aprovação de crédito e assim testar se os modelos têm boa aderência e boa capacidade de previsão comportamental foi dividida a amostra em 2 conjuntos mutuamente exclusivos onde se incluem os dados que foram utilizados para conferirem a generalização do modelo e a medida do seu desempenho uma vez que, de acordo com Hair *et al.* (2005), na regressão logística, para se obter uma boa estimativa da eficiência de classificação do modelo, a amostra deve ser separada em duas partes uma para construção do modelo e a outra para validação do modelo.

Aquando da construção de um modelo é importante testar não só a sua confiança mas também a sua capacidade de generalização. Seria demasiado simplista esperar que um modelo ao ter sido criado pela selecção de uma amostra de dados e ao demonstrar que assegura uma boa previsão para esse conjunto, o faça igualmente para a restante população, (Pallant, 2006). Para avaliar se o modelo mantém o seu poder preditivo para outras amostras providas da mesma população, são necessários testes para a sua validação. Segundo Hair *et al.* (2005), a validação do modelo é especialmente importante quando ele é usado com a finalidade de previsão de resultados.

Assim, neste estudo, os 1000 casos foram depois divididos em 2 sub-amostras provenientes do mesmo universo de interesse. A separação deu-se de modo aleatório com 70% dos dados para a construção do modelo e 30% para sua previsão. O objectivo desta separação foi garantir que os dados não fossem os mesmos para a construção do modelo, e para a validação e desse modo ser possível estabelecer a validação. Os clientes seleccionados através do processo de amostragem aleatória tiveram os seus dados processados no *software SPSS* versão 13.0.

4. RESULTADOS E DISCUSSÃO DOS MODELOS

4.1 CARACTERIZAÇÃO DA AMOSTRA

No presente estudo, foi utilizada uma amostra de 1000 indivíduos a quem uma instituição financeira concedeu crédito. Destes, 500 representavam os bons clientes e os outros 500, os maus clientes. Tal como ilustra o Gráfico 4.1, cerca de 55% destes clientes eram do sexo masculino e os restantes 45% do sexo feminino.

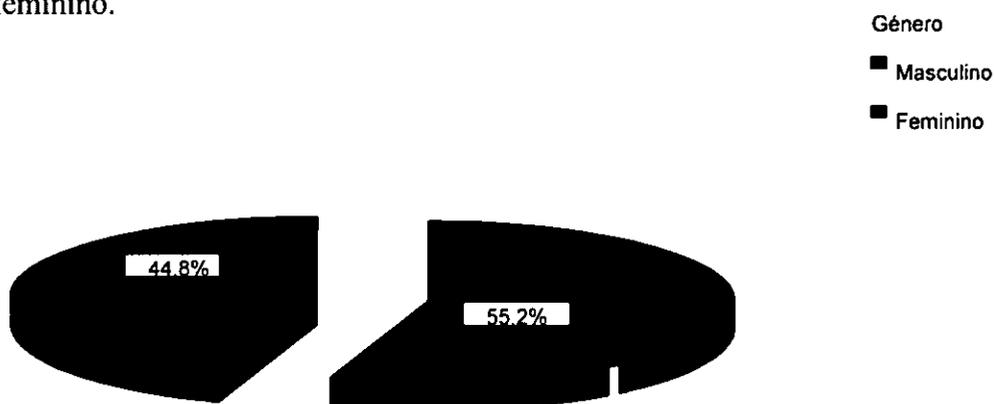


Gráfico 4.1: Distribuição dos clientes que compuseram a amostra com base no género

Tendo em conta o estado civil dos clientes tanto do sexo feminino quanto do sexo masculino, a maioria (cerca de 58 %) dos clientes não são casados. Dos clientes do sexo masculino, apenas 41.5% destes é que são casados e quanto ao sexo feminino 42 % dos clientes não são casados. Assim sendo, dos 1000 clientes que compõem a amostra, 58.3 % destes não são casados e somente 41.7 % é que são casados (ver Tabela 4.1).

		Género				Total	
		Masculino		Feminino			
		N	%	N	%	N	%
Estado Civil	não casado	323	58.5	260	58.0	583	58.3
	casado	229	41.5	188	42.0	417	41.7
Total		552	100.0	448	100.0	1000	100.0

Tabela 4.1: Caracterização da amostra de acordo com o estado civil e o género

Quando se toma em consideração a idade verifica-se que a maior percentagem (48.2%) dos clientes envolvidos no estudo tinham idade entre os 35 e os 50 anos (Gráfico 4.2), e apenas 14.4% tinham mais de 50 anos. Os restantes 37.4% correspondem aos clientes com menos de 35 anos.

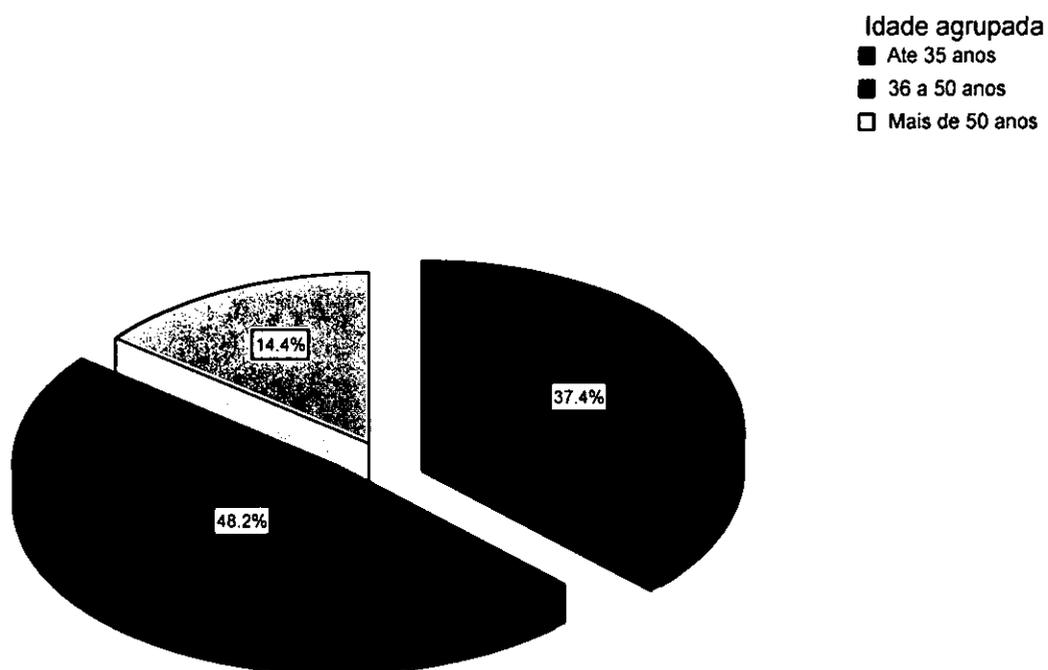


Gráfico 4.2: Distribuição dos clientes que compuseram a amostra de acordo com o grupo etário

Tendo como base o sector profissional construiu-se a Tabela 4.2. Nela pode-se verificar que cerca de 26% dos clientes que compuseram a amostra do estudo não indicaram a sua profissão e a maior percentagem exerce a sua profissão no sector do comércio (38.5). Apenas 5.1% não exerce qualquer actividade profissional, menos de 1% tem a sua ocupação no sector produtivo da agricultura, silvicultura, pesca e extracção mineira, 3.7% na indústria manufactureira, 5.3% no sector de transportes e comunicações, 1% estudantes, 4.7% profissionais da saúde ou educação, 9.3% têm a sua ocupação no sector de administração e 6% são artesãos.

	Frequência	Percentagem (%)
não indicada	260	26.0
agricultura, silvicultura, pesca, extracção mineira	4	0.4
indústria manufactureira	37	3.7
transportes, comunicações	53	5.3
sem ocupação	51	5.1
comércio	385	38.5
estudante	10	1.0
educação, saúde	47	4.7
administração	93	9.3
artesanato	60	6.0
Total	1000	100.0

Tabela 4.2: Distribuição dos clientes que compuseram a amostra, por Sector Profissional

Quanto ao nível de formação académica dos clientes que compuseram a amostra cuja distribuição é apresentada na Tabela 4.3, constatou-se que cerca de 32 % dos clientes não forneceram informação, cerca de 2% afirmaram ser analfabeto, e percentagem idêntica referiu ter formação técnica de nível básico. Do total, 18.0 % referiu ter formação técnica de nível médio, valor semelhante ao do que referiu ter o ensino secundário geral. Apenas 6.8% dos clientes referiram ter o nível superior de formação académica.

	Frequência	Percentagem (%)
Não indicado	<i>315</i>	<i>31.5</i>
Analfabeto	<i>23</i>	<i>2.3</i>
Ensino Primário	<i>211</i>	<i>21.1</i>
Ensino secundário	<i>181</i>	<i>18.1</i>
Técnico Básico	<i>22</i>	<i>2.2</i>
Técnico Médio	<i>180</i>	<i>18.0</i>
Superior	<i>68</i>	<i>6.8</i>
Total	<i>1000</i>	<i>100.0</i>

Tabela 4.3: Distribuição dos clientes que compuseram a amostra de acordo com o nível de formação acadêmica

Quanto à Agência, como local onde foi obtido o crédito, verifica-se pelos resultados apresentados na Tabela 4.4 que de entre as 6 agências que tiveram os seus clientes incluídos no estudo, a Agência 1 foi a mais representada com cerca de 55% dos clientes, seguida pela Agência 2 (29.2%), Agência 3 (10.8%). As restantes 3 Agências foram representadas por menos de 5%.

	Frequência	Percentagem (%)
1	<i>553</i>	<i>55.3</i>
2	<i>292</i>	<i>29.2</i>
3	<i>108</i>	<i>10.8</i>
4	<i>29</i>	<i>2.9</i>
5	<i>16</i>	<i>1.6</i>
6	<i>2</i>	<i>0.2</i>
Total	<i>1000</i>	<i>100.0</i>

Tabela 4.4: Distribuição dos clientes que compuseram a amostra, por Agência

Tal como pode ser constatado na Tabela 4.5, o valor do crédito concedido variou entre um mínimo de 3000.00 meticais e um máximo de 120000.00 meticais. O valor médio concedido foi de 12587.93 meticais, com um desvio padrão de 10818.08 meticais, sendo o valor concedido com mais frequência o de 20000.00 meticais.

	N	Mínimo	Máximo	Média	Moda	Desvio Padrão
Valor do crédito concedido	1000	3000	120000	12587.93	20000	10818.087

Tabela 4.5: Estatística descritiva relativa ao valor do crédito concedido

4.2 APRESENTAÇÃO DOS MODELOS

Nas secções que se seguem são apresentados, analisados e discutidos os modelos de *credit scoring* para a análise de risco de crédito e os testes a estes associados pela aplicação da regressão logística.

4.2.1 MODELO DE APROVAÇÃO DE CRÉDITO (*APPLICATION SCORING*)

O modelo de aprovação de crédito estimado por meio da técnica estatística de regressão logística foi o seguinte:

	B	S.E.	Wald	df	p	Exp(B)
Idadeagrupo (1)	1.135	0.486	5.453	1	0.020	3.112
Prof_agr(8)	-1.253	0.692	3.279	1	0.040	0.901
Género(1)	0.704	0.254	7.688	1	0.006	2.022
Constante	-2.931	0.806	13.235	1	0.000	0.053

Tabela 4.6: Estimativas dos parâmetros do modelo de aprovação de crédito

A análise da Tabela 4.6 sugere que, neste modelo de aprovação de crédito, a *idadeagrup(1)* é a característica que exerce maior efeito na propensão à inadimplência. Por outro lado, a variável que contribuiu para a redução da inadimplência, isto é, a variável que diminui a probabilidade do cliente ser inadimplente (mau pagador), é a *prof-agr(8)*.

O formato da equação deste modelo pode ser escrito como:

$$P_i = \frac{1}{1 + e^{-z_i}}$$

$$Z_i = 2,931 + 1,135 \text{Idadeagrup}(1) - 1,253 \text{Prof-agrup}(8) + 0,704 \text{género}(1)$$

Onde:

Idadeagrup (1)= Idade entre os 36 e os 50 anos;

Prof-agrup (8)= Sector profissional-administração

Género (1)= Género feminino;

Z_i = Score do cliente *i*;

P_i = Probabilidade de inadimplência do cliente *i*.

De acordo com o modelo acima, as variáveis que mais discriminam os bons dos maus clientes são: a idade, a profissão e o género. Assim, ao realizar uma classificação de um solicitante com base no modelo construído, por exemplo, o facto dele pertencer ao grupo etário dos 36 aos 50 anos irá contribuir significativamente para a sua classificação como possível cliente inadimplente. Requerentes de crédito com as características referentes às variáveis explicativas do modelo demandam, portanto, uma análise mais detalhada e cuidadosa, com maior zelo por parte do analista de crédito, já que têm considerável propensão à inadimplência. Esse efeito de cada variável

explicativa do modelo sobre os scores da operação de crédito pode ser descrito a partir da análise dos coeficientes considerando que:

- *Idadeagrup(1)*- o sinal positivo desta variável significa que indivíduos que se encontram na faixa etária dos 36 aos 50 anos estão mais propensos a serem inadimplentes, isto é, maus pagadores. Este resultado, sugere que os clientes nesta faixa etária (dos 36 anos aos 50 anos) possam ter maior probabilidade de inadimplência do que os indivíduos pertencentes a outras faixas etárias (até 35 anos e mais de 50 anos). Sendo que as chances de um indivíduo pertencente a esta faixa etária seja 3.112 vezes maior do que aqueles que pertencem a outra faixa. Uma possível explicação para tal tendência pode ser o facto de que indivíduos pertencentes a esta faixa etária procurarem concretizar um maior número de projectos relacionados com a sua assumpção como adultos.
- *Prof-agrup(8)*- para esta variável o sinal negativo significa uma menor propensão para a inadimplência só que neste caso, pelos indivíduos que exercem a sua actividade profissional no sector da administração, sendo a chance de se tornar inadimplente de 1.11 vezes menor do que os indivíduos de outros sectores. Uma possível explicação para tal propensão pode ser o facto de que os funcionários do sector da administração serem mais hábeis a administrar os seus próprios bens dada a experiência e as características referentes à sua actividade profissional.
- *Género (1)*: o sinal positivo do coeficiente dessa variável significa que clientes pertencentes ao género feminino têm maior propensão à inadimplência que os do sexo masculino, sendo a chance de se tornar inadimplente 2.022 vezes maior do que os indivíduos do sexo masculino. Provavelmente devido ao facto de que tradicionalmente são os indivíduos do sexo masculino que se responsabilizam pela gestão e administração a nível familiar reduzindo a possibilidade dos indivíduos do outro grupo desenvolverem habilidades que lhe permitiriam gerir os empréstimos com menos propensão à inadimplência.

Para obter a significância de cada coeficiente logístico (β), em regressão logística, fez-se uso do teste de *Wald*. Os resultados deste teste são os apresentados na Tabela 4.7. Assim sendo, a um nível

de significância de 95% não há evidências suficientes para rejeitar a hipótese nula, isto, é os coeficientes logísticos são diferentes de zero, representando uma adequação do modelo para explicar a variável dependente “situação do cliente”.

		χ^2	df	p
Passo 1	Passo	369.194	6	0.000
	Bloco	369.194	6	0.000
	Modelo	369.194	6	0.000
Passo 2	Passo	88.698	9	0.000
	Bloco	457.893	15	0.000
	Modelo	457.893	15	0.000
Passo 3	Passo	13.330	2	0.001
	Bloco	471.223	17	0.000
	Modelo	471.223	17	0.000
Passo 4	Passo	8.057	1	0.005
	Bloco	479.279	18	0.000
	Modelo	479.279	18	0.000

Tabela 4.7: Estatística *GM* do modelo de aprovação de crédito

Pela Tabela 4.7, pode-se concluir que as variáveis foram seleccionadas em 4 passos e o resultado da aplicação do teste GM, para avaliar a significância das variáveis explicativas incluídas no modelo, teve como resultado $\chi^2 = 479.279$; $p = 0.000 (<0.05)$ o que portanto leva à rejeição da hipótese nula de que as variáveis independentes não explicam a variável dependente “situação do cliente”, pois pelo menos um dos coeficientes é estatisticamente diferente de zero.

Passo	-2 Log-Likelihood	Cox-Snell R²	Nagelkerke R²
1	196.230	0.346	0.461
2	165.302	0.438	0.584
3	132.088	0.522	0.696
4	127.474	0.533	0.711

Tabela 4.8: Medidas *pseudo R²* do modelo de aprovação de crédito

A partir dos valores apresentados na Tabela 4.8, pode ser verificado que a medida que foram sendo introduzidas as variáveis no modelo, o valor de $-2 \text{ Log-Likelihood}$ foi reduzindo de 196.230 para 127.474, o que indica um bom ajustamento geral do modelo pelo baixo valor encontrado. No modelo gerado, obteve-se como valor de R^2 Cox & Snell, variando de 0.346 para 0.533 e de R^2 Nagelkerke, variando de 0.461 para 0.711. Assim sendo, de acordo com o indicador de Cox & Snell 53.3% das variações registadas na variável dependente são explicadas pelo conjunto de variáveis independentes e com o de Nagelkerke 71.1%, traduzindo-se assim que o modelo tem um bom poder de discriminação.

Adicionalmente, o *pseudo* R^2 para o modelo *logit* foi de 0.3504, o que comprova mais uma vez que o modelo proposto discrimina os bons dos maus pagadores de uma maneira satisfatória, assim calculado:

$$R^2 = \frac{2LL_{nulo} - (-2LL_{modelo})}{-2LL_{nulo}} = \frac{196.230 - 127.474}{196.230} = 0.3504$$

O teste de Hosmer-Lemeshow é apresentado nas tabelas que se seguem (Tabelas 4.9 e 4.10):

		situação = Bom cliente		situação = Mau cliente		Total
		Observado	Esperado	Observado	Esperado	
Passo 4	1	66	66.584	6	5.416	72
	2	57	58.017	11	9.983	68
	3	57	52.477	9	13.523	66
	4	44	48.683	20	15.317	64
	5	46	42.546	14	17.454	60
	6	37	42.620	31	25.380	68
	7	36	32.073	34	37.927	70
	8	0	0.000	73	73.000	73
	9	0	0.000	60	60.000	60
	10	0	0.000	85	85.000	85

Tabela 4.9: Tabela de contingência do teste Hosmer- Lemeshow do modelo de aprovação de crédito

		χ^2	df	p
Passo	1	0.000	3	1.000
	2	1.092	6	0.982
	3	10.596	8	0.226
	4	7.811	8	0.452

Tabela 4.10: Teste de Hosmer-Lemeshow do modelo de aprovação de crédito

O resultado obtido pelo teste de Hosmer-Lemeshow apresentado na Tabela 4.10 com um χ^2 ($df=8$) = 7.811 e um valor de $p = 0.452$, levando à não rejeição da hipótese de que não existem diferenças significativas entre as classificações observadas e as classificações efectuadas pelo modelo ($p < 0.05$), o que configura, mais uma vez, uma boa aderência do modelo à realidade observada. Assim sendo, o modelo ajusta-se aos dados sendo capaz de produzir estimativas fiáveis.

4.2.2 MODELO COMPORTAMENTAL (*BEHAVIOURAL SCORING*)

Para a construção do modelo comportamental de concessão de crédito foi necessário acrescentar, para além das variáveis seleccionadas para construção do modelo de aprovação de crédito, as variáveis referentes ao histórico do cliente com a instituição, como é o caso da variável número de créditos anteriores com a instituição e a situação do cliente em créditos anteriores, de modo a ter informações que possam auxiliar o analista de crédito na gestão dos créditos de solicitantes que já são clientes da instituição. A Tabela 4.11 apresenta o modelo comportamental de crédito:

	B	S.E.	Wald	df	p	Exp(B)
Idadeagr(1)	1.061	0.495	4.598	1	0.032	2.891
Prof_agr(8)	-1.137	0.720	2.496	1	0.014	0.518
Género(1)	0.753	0.260	8.379	1	0.004	2.124
Ncredit	-0.731	0.178	16.806	1	0.000	0.481
Constante	-1.813	0.870	4.341	1	0.037	0.163

Tabela 4.11: Estimação dos parâmetros do modelo comportamental de crédito (*behavioural scoring*)

O modelo estimado foi o seguinte:

$$P_i = \frac{1}{1 + e^{-z_i}}$$

$$Z_i = 1,813 + 1,061\text{Idadeagrup}(1) - 1,137\text{Prof-agrup}(8) + 0,753\text{g\u00e9nero}(1) - 0,731\text{ncredit}$$

Onde:

Idadeagrup (1)= idade entre os 36 e os 50 anos;

Prof-agrup (8)= sector profissional-administra\u00e7\u00e3o;

G\u00e9nero (1)= g\u00e9nero feminino;

ncredit= n\u00famero de cr\u00e9ditos anteriores com a institui\u00e7\u00e3o;

Z_i = \u00e9 o score do cliente i ;

P_i = probabilidade de inadimpl\u00eancia do cliente i .

Para al\u00e9m das vari\u00e1veis que discriminam os adimplentes dos clientes inadimplentes do modelo de aprova\u00e7\u00e3o de cr\u00e9dito (idadeagr (1); prof_agr (8) e g\u00e9nero), o modelo *behavioural scoring* incorporou a vari\u00e1vel n\u00famero de cr\u00e9dito anteriores com a institui\u00e7\u00e3o (ncredit). Os coeficientes das vari\u00e1veis que j\u00e1 foram consideradas no modelo de aprova\u00e7\u00e3o apresentaram os mesmos sinais, e, portanto, efeito similar sobre os scores de cada cliente, o que indica n\u00e3o haver incoer\u00eancia nessas vari\u00e1veis entre os dois modelos. O efeito da vari\u00e1vel n\u00famero de cr\u00e9ditos anteriores (ncredit) sobre a vari\u00e1vel dependente \u00e9 analisado a seguir.

- N\u00famero de cr\u00e9ditos anteriores com a institui\u00e7\u00e3o: o sinal negativo do coeficiente mostra que esta vari\u00e1vel contribui bastante para a redu\u00e7\u00e3o da inadimpl\u00eancia. Assim sendo, clientes com maior n\u00famero de cr\u00e9ditos anteriores tendem a ser menos inadimplentes.

Em relação à significância estatística das variáveis explicativas do modelo comportamental de crédito e com base nos resultados do teste Wald, rejeita-se a hipótese das variáveis serem iguais a zero, isto é, todos os coeficientes foram significativos. Assim sendo, comprova-se que o modelo na sua forma geral foi significativo (Tabela 4.11).

		χ^2	df	p.
Passo1	Passo	369.194	6	0.000
	Bloco	369.194	6	0.000
	Modelo	369.194	6	0.000
Passo 2	Passo	88.698	9	0.000
	Bloco	457.893	15	0.000
	Modelo	457.893	15	0.000
Passo3	Passo	22.067	1	0.000
	Bloco	479.960	16	0.000
	Modelo	479.960	16	0.000
Passo4	Passo	8.919	1	0.003
	Bloco	488.879	17	0.000
	Modelo	488.879	17	0.000
Passo5	Passo	9.946	2	0.007
	Bloco	498.825	19	0.000
	Modelo	498.825	19	0.000

Tabela 4.12: Estatística GM do modelo comportamental de crédito

Na Tabela 4.12, pode-se verificar que as variáveis independentes explicam o modelo ($\chi^2 = 498.825$; $p = 0.000 < 0.05$), isto é, a nível de significância de 95%, há evidências suficientes para rejeitar a hipótese das variáveis independentes não explicarem a variável dependente (situação do cliente).

Na Tabela 4.13 são apresentadas as medidas de “pseudo” R^2 . De acordo com o indicador de Cox & Snell 51.7% das variações registadas na variável dependente são explicadas pelo conjunto de variáveis independentes e com o de Nagelkerke 68.9%, o “pseudo” R^2 para o modelo *logit* foi de 0.2228, calculado com base na fórmula a seguir apresentada. Todos os valores apresentados são satisfatórios, de acordo com os critérios referidos no Capítulo 3.

$$R^2 = \frac{2LLnulo - (-2LLmodelo)}{-2LLnulo} = \frac{581.804 - 452.173}{581.804} = 0.2228$$

		-2 Log-likelihood	Cox & Snell R ²	Nagelkerke R ²
Passo	1	581.804	.416	.555
	2	493.105	.487	.649
	3	471.038	.503	.671
	4	462.119	.510	.680
	5	452.173	.517	.689

Tabela 4.13: Medidas pseudo R² do modelo comportamental de crédito

O teste de Hosmer-Lemeshow é apresentado nas Tabelas 4.14 e 4.15. O resultado obtido por este teste confirma o bom ajustamento do modelo aos dados, isto é, não há evidências suficientes para rejeitar a hipótese de que os dados ajustam-se à realidade observada (χ^2 (df=8, N=1962.09) = 7.275, $p= 0.507$).

		Situação do cliente = Bom		Situação do cliente = Mau		Total
		Observado	Esperado	Observado	Esperado	
Passo 5	1	68	68.353	4	3.647	72
	2	57	60.571	12	8.429	69
	3	61	56.715	8	12.285	69
	4	58	54.991	15	18.009	73
	5	50	50.977	25	24.023	75
	6	35	37.729	32	29.271	67
	7	14	13.664	56	56.336	70
	8	0	0.000	72	72.000	72
	9	0	0.000	34	34.000	34
	10	0	0.000	85	85.000	85

Tabela 4.14: Tabela de contingência do teste Hosmer- Lemeshow do modelo comportamental de crédito

		χ^2	<i>df</i>	<i>p</i>
Passo	1	<i>0.000</i>	<i>3</i>	<i>1.000</i>
	2	<i>1.092</i>	<i>6</i>	<i>0.982</i>
	3	<i>5.470</i>	<i>8</i>	<i>0.706</i>
	4	<i>8.741</i>	<i>8</i>	<i>0.365</i>
	5	<i>4.765</i>	<i>8</i>	<i>0.782</i>

Tabela 4.15: Teste de Hosmer-Lemeshow do modelo comportamental de crédito

4.3 VALIDAÇÃO DOS MODELOS

Após o desenvolvimento dos modelos de aprovação de crédito e comportamental avaliou-se a sua capacidade preditiva através da Matriz de Confusão ou Tabela de Classificação e da construção de uma Curva ROC.

4.3.1 MATRIZ DE CONFUSÃO OU TABELA DE CLASSIFICAÇÃO

Construiu-se uma matriz de classificação de forma a identificar a quantidade de clientes classificados correcta e incorrectamente pelos modelos. Nas Tabelas 4.16 e 4.17 são apresentadas as percentagens de acertos dos modelos desenvolvidos.

		Classificação do modelo	
		Bom pagador	Mau pagador
Classificação da amostra	Bom pagador	<i>88.2 %</i>	<i>11.8 %</i>
	Mau pagador	<i>26.5 %</i>	<i>73.5 %</i>
Percentagem geral		<i>80.9 %</i>	

Tabela 4.16: Matriz de classificação do modelo de aprovação de crédito (% acertos)

No modelo de aprovação de crédito, dos bons clientes, 80.2 % dos clientes foram classificados correctamente e 11.8% dos clientes foram classificados como maus pagadores enquanto foram bons clientes, tendo portanto sido mal classificados. Em relação ao grupo dos maus pagadores para o mesmo modelo (aprovação), 75 clientes foram classificados correctamente e 27 clientes foram classificados como bons enquanto foram maus clientes, o que corresponde a uma taxa de acerto de 73.5%. Assim sendo, a taxa de acerto geral do modelo foi de 80.9%.

		Classificação do modelo	
		Bom pagador	Mau pagador
Classificação da amostra	Bom pagador	89.2 %	11.8 %
	Mau pagador	9.6 %	80.4 %
Percentagem geral		80.9 %	

Tabela 4.17: Matriz de classificação do modelo comportamental de crédito (% de acertos)

Na Tabela 4.17 pode-se observar que dos bons pagadores, 88.2% foram correctamente classificados enquanto 11.8% foram classificados como maus enquanto são bons pagadores. Para o caso dos maus clientes, 80.4 % dos clientes foram correctamente classificados e apenas 9.6% dos clientes foram classificados como bons enquanto foram maus pagadores.

De um modo geral, o modelo comportamental (*behavioural scoring*) classifica correctamente 84.9% dos clientes. Sendo assim, os modelos explicam melhor os bons pagadores o que pode ser justificado pelo facto de no universo dos clientes da instituição financeira, usada como local de estudo, ser mais frequente encontrar bons pagadores do que maus pagadores. Provavelmente estes resultados revelam também uma boa qualidade do sistema de análise de risco de crédito em uso na instituição.

4.3.2 CURVA DE ROC

A área sob a curva mede a capacidade de discriminação do modelo. A Curva ROC do modelo de aprovação de crédito, representada nos Gráficos 4.3 e 4.4, revelam que a área sob a curva é de 0.931 e 0.954 respectivamente.

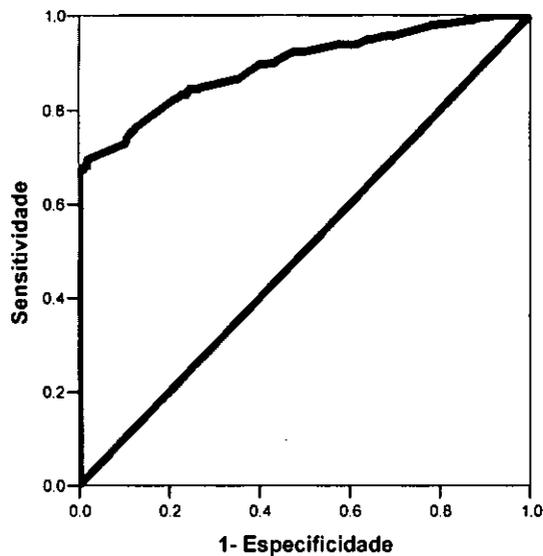


Gráfico 4.3: Curva de ROC do modelo de aprovação de crédito

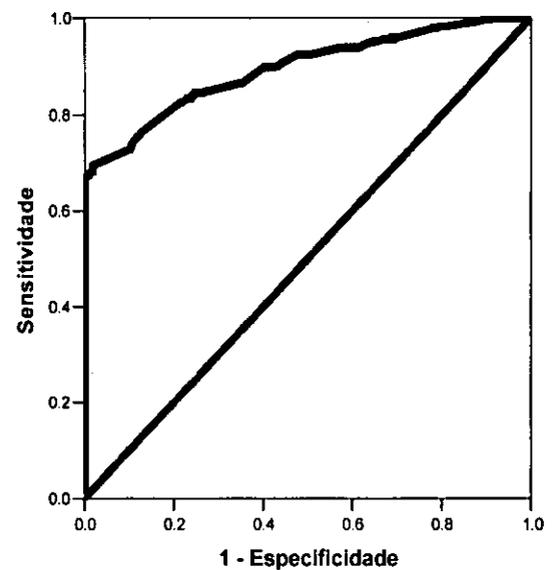


Gráfico 4.4: Curva de ROC do modelo comportamental de crédito

Segundo a classificação proposta por Marroco (2007), descrita no Capítulo 2, os valores obtidos pela Curva de ROC indicam um excepcional poder de discriminação dos modelos. Deste modo, os modelos apresentados são válidos para discriminar os bons dos maus pagadores.

5. CONCLUSÕES E RECOMENDAÇÕES PARA A ANÁLISE DE RISCO DE CRÉDITO

5.1 CONCLUSÕES

Neste trabalho foram desenvolvidos modelos quantitativos de *credit scoring* de aplicação de crédito (*application scoring*) e comportamental (*behavioural scoring*) para a avaliação do risco de crédito numa instituição financeira usando a regressão logística com o objectivo de apresentar ferramentas que possam auxiliar a identificar e prever que clientes poderão ser (ou não) inadimplentes

O modelo de aprovação agregou 3 variáveis (“profissão”, “género” e “idade”), que no modelo comportamental foram acrescidas pela variável “número de créditos anteriores”. Os resultados do estudo mostraram que os clientes pertencentes ao grupo etário entre os 18 e os 35 anos e indivíduos do sexo feminino, apresentavam uma maior propensão para a inadimplência enquanto que os profissionais do sector de administração apresentavam uma menor propensão para a inadimplência. Estes resultados são coerentes com o que se observa na realidade.

Os modelos gerados apresentaram um resultado muito bom, com um forte suporte empírico tanto na amostra para geração do modelo quanto na de validação. Com eles, puderam ser identificados os clientes enquadrados como bons pagadores (adimplentes) e maus pagadores (inadimplentes), bem como a probabilidade de risco de crédito associada a eles, com uma percentagem de acerto superior a 80%.

Embora estes modelos mostraram-se melhores na classificação correcta dos clientes adimplentes do que dos inadimplentes é possível a utilização de modelos de *credit scoring* na instituição usada como local de estudo, como instrumentos de apoio ao processo de avaliação do risco de crédito. No entanto, é necessário ressaltar que o uso de modelos *credit scoring* não substituem a avaliação feita pelo analista de crédito, estes modelos complementam a avaliação feita pelos analistas de crédito e por isso sugere-se que sejam agregados aos modelos de análise de risco de crédito em uso na instituição. Com o uso destes modelos será possível a identificação prévia da probabilidade de ocorrência de inadimplência. Desta forma, o processo de avaliação do risco tornar-se-á, mais preciso, criterioso, com menos ambiguidade e um menor grau de subjectividade. Isso, sem dúvida, contribui na a redução da inadimplência na instituição financeira e consequentemente para a redução dos custos operacionais.

5.2 POSSÍVEIS CONTRIBUIÇÕES PARA A ANÁLISE DE RISCO DE CRÉDITO

A utilização das técnicas estatísticas na análise do risco de crédito tem sido cada vez maior. Por isso, na elaboração de trabalhos futuros poderão ser utilizadas outras técnicas estatísticas, tais como redes neurais e árvores de decisão para obter modelos de risco de crédito. Para além disso estudos futuros deverão basear-se em dados mais contemporâneos uma vez que os modelos de *credit scoring* devem ser actualizados de forma permanente de modo a não ter degradado o seu poder preditivo.

No entanto, qualquer que seja a técnica a aplicar no desenvolvimento de modelos de avaliação de crédito deve ser tomada em consideração a qualidade de preenchimento dos dados pois será de todo necessário que se adotem medidas de maior rigor no controle do preenchimento da ficha de solicitação e se procure que o lançamento de dados seja o mais fiel e completo possível. Neste contexto sugere-se que se ofereça aos operadores do sistema a possibilidade de treinamento contribuindo-se assim para que a instituição possa melhor fazer uso da informação relativa aos solicitantes que habitualmente já é solicitada.



BIBLIOGRAFIA

1. Akiama, S. R. (2008). *Probabilidade de inadimplência de grandes empresas no sistema financeiro nacional*. Tese de Mestrado em Ciências Contábeis. Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo. Disponível em www.teses.usp.br/teses/disponiveis/12/12136/tde-04062008-120550
2. Araújo, E. A. (2006). *Modelagem de risco de crédito: aplicação de modelos credit scoring no fundo rotativo de ação da cidadania-cred cidadania*. Tese de Mestrado em Administração, Universidade Federal de Pernambuco, Recife. Disponível em www.bdttd.ufpe.br/tedeSimplificado/tde_buscaarquivo.php
3. Boaventura, E. M. (2004). *Metodologia da Pesquisa: monografia, dissertação, tese*. 1 Ed., São Paulo: Atlas.
4. Carpenter, E. M. L. (2006). *Um modelo de análise de risco de crédito de clientes em relações B2B*. Tese de Mestrado em Administração de empresas, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro. Disponível em www.cipedya.com/doc/161519.
5. Carvalho, A. T. (2004). *Modelo de previsão de inadimplência para empresas comerciais*. Tese de Mestrado em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis. Disponível em <http://teses.eps.ufsc.br/defesa/pdf/9939.pdf>
6. Chorão, L. A. R. (2005). *Credit Scoring: Logit vs Redes Neurais um exemplo aplicativo a cartões de crédito*. Tese de Mestrado em Estatística, Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Lisboa.

7. Chorão, L. A. R. e Martins, M. R. O. (s/d). *Unidade de aprendizagem. Módulo XI- Credit Scoring*. Acedido a 30 de Setembro, 2008.
8. Costa, R. D. (2007). *Um modelo de previsão de insolvência para bancos privados nacionais*. Trabalho de Licenciatura de Matemática Aplicada a Negócios. Universidade de São Paulo, Ribeirão Preto. Disponível em www.ffclrp.usp.br/dfm/tccman/2008/Costa_RDM.pdf
9. Da Cunha, J. V. A. ; Martins, G. A. ; Júnior, E. B. C. (s/d). *Uma aplicação da regressão logística no inventário de estilos de aprendizagem de Carfield (LSI) sob a ótica das reprovações acadêmicas*. Acedido a 27 de Dezembro, 2008. Disponível em www.congressoeac.locaweb.com.br/artigos62006/224.pdf
10. Gil, A. C. (1995). *Métodos e técnicas de pesquisa social*. 4ed., São Paulo: Atlas.
11. Gonçalves, E. B. (2005). *Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéricos*. Tese de Mestrado em Administração, Universidade de São Paulo, São Paulo. Disponível em www.teses.usp.br/teses/disponiveis/12/12139/tde-09042008/144032/.
12. Guimarães, A. L. De Souza (2008). *Avaliando a classificação de risco de crédito em operações indiretas com garantia do fundo de aval*, 15 (30), 39-61. Disponível em www.bndes.gov.br/conhecimento/revista/rev3002.pdf
13. Gujarati, D. N. (2004). *Econometria Básica*, 4ª Ed., São Paulo: Markron books.
14. Hair, J. F.; Anderson, R. E.; Tatham, R. L.; Black, W. C. (2005). *Análise Multivariada de Dados*. 5ª Ed., Porto Alegre : Bookman.
15. Lima, L. R. S. (2008). *Uma proposta de construção de um modelo de avaliação de risco de crédito para micros e pequenas empresas financiadas pela Desenhahia*. Acedido a 22 de Novembro, 2002. Disponível em www.desenhahia.ba.gov.br/estudos/cadernos.asp
16. Marroco, J. (2007). *Análise Estatística: com utilização do SPSS*, 3ª Ed., Lisboa: Sílabo.

17. Minussi, J. A. ; Damacena, C. e Júnior, W. L. N. (2002). *Um modelo de previsão de solvência utilizando regressão logística*, 6 (3), 109-128. Disponível em www.anpad.org/rac/vol_06/htm/rac-v6-n3-jam.pdf
18. Neto, A. (2001). *Mercado financeiro*. 4ª Ed., São Paulo: Atlas.
19. Neto, A. A. e Brito, G. A. S. (s/d). *Modelo de classificação de risco de crédito de empresas*. Universidade de São Paulo, 19, 18-49. Acedido a 11 de Junho, 2008. Disponível em www.eac.fea.usp.br/cadernos/completos/46/giovani-alexandre.pdf
20. Neto, A. A. e Carmona, C. (2007). *Modelagem do risco de crédito: Um estudo do segmento de pessoas físicas em um banco de varejo*. Núcleo de estudos em finanças e investimentos do programa de Pós-graduação em Administração da UFPE. Disponível em www.dca.ufpe.br/nefi
21. Pallant, J. (2006). *SPSS Survival Manual*. 2nd Ed. Bookhouse, Sydney Australia
22. Paula, G. A. (2004). *Modelos de regressão com apoio computacional*. Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Disponível em www.feferraz.net/br/Livros/Modelos_de_regressao_com_apoio_computacional
23. Penha, R. N. (2002). *Um estudo sobre regressão logística binária*. Universidade Federal de Itajubá, Trabalho de Licenciatura em Engenharia de Produção, Minas Gerais. Disponível em www.iem.efei.br/dpr/td/producao2002/pdf/renata.pdf
24. Pereira, G. H. De A. (2004). *Modelos de risco de crédito de clientes: uma aplicação a dados reais*. Tese de Mestrado em Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Disponível em www.teses.usp.br/teses/disponiveis/45/45133/tde-28122004-224257

25. Raymundo, P. J. (2002). *Fatores considerados pelas instituições financeiras para a determinação do risco do cliente e do limite de crédito para capital de giro das micro e pequenas empresas*. Tese de Mestrado em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis. Disponível em www.teses.eps.ufsc.br/defesa/pdf/10182.pdf.
26. Régis, D. E. (2007). *Aplicação do modelo multi-estado de Markov em cartões de crédito*. Tese de Mestrado em Economia, Faculdade IBMEC São Paulo, São Paulo. Disponível em www.ibmecsp.edu.br
27. Rezende, F. C (2007). *Construção de modelos de classificação de risco de crédito para empresas brasileiras com base em indicadores contábeis*. Tese de Mestrado em Economia e Finanças, IBMEC São Paulo, Faculdade de Economia e Administração. Disponível em http://tede.ibmecsp.edu.br/tde_busca/arquivo.php
28. Silva, E. R. (2006). *Aplicação de metodologia de dados em painel em modelos de behaviour score do varejo*. Tese de Mestrado em Finanças e Macroeconomia Aplicada, IBMEC, São Paulo. Disponível em www.ibmecsp.edu.br
29. Souza, L. D.; Bastos, C. G.; Costa, M. L. & Pinto, R. M. C. (2004). *Avaliação do risco de crédito: Uma aplicação estatística no segmento Telecom*. Universidade Federal de Uberlândia – Faculdade de Matemática – Uberlândia, Minas Gerais. Disponível em www.posgraduação.ufla.br/gauss/congress/49rbras/pub/t095.pdf.

ANEXO

Tabela A-1: Variáveis consideradas na análise e a sua natureza

Variável	Atributos/ descrição	Natureza/Formato da Variável
Gênero	Masculino Feminino	Qualitativa (categorias)
Idade	Até aos 35 anos De 36 anos aos 50 anos Mais de 50 anos	Quantitativa (em anos)
Estado civil	Não Casado Casado	Qualitativa (categorias)
Nível acadêmico	Não indicado Analfabeto Ensino Primário Ensino Secundário Técnico Básico Técnico Médio Nível Superior	Qualitativa (categorias)
Sector Profissional	Não indicada Agricultura, silvicultura, pesca, extracção mineira Indústria manufactureira Transportes, comunicações Sem ocupação Comércio Estudante Educação, saúde Administração Artesãos	Qualitativa (categorias)
Agência	1 2 3 4 5 6	Qualitativa (categorias)
Valor do empréstimo		Quantitativa (em meticais)
Número de créditos anteriores		Quantitativa (em números inteiros)
Histórico com a instituição	Não possui crédito anterior Não possui atraso anterior (30 dias ou mais) Possui atraso anterior superior a 30 dias	Qualitativa (categorias)

Situação do cliente	Bom cliente	Qualitativa (categorias)
	Mau cliente	

Tabela A-2: Variáveis dummies

Variáveis	Categorias	Variável dummy									
		D _g	D _{i1}	D _{i2}	D _{ec}	D _{nac1}	D _{nac2}	D _{nac3}	D _{nac4}	D _{nac5}	D _{nac6}
Gênero	Masculino	1									
	Feminino	0									
Idade	Até aos 35 anos		1	0							
	Dos 35 aos 50 anos		0	1							
	Mais de 50anos		0	0							
Estado Civil	Não casado				1						
	Casado				0						
Nível acadêmico	Não indicado					1	0	0	0	0	0
	Analfabeto					0	1	0	0	0	0
	Ensino Primário					0	0	1	0	0	0
	Ensino Secundário					0	0	0	1	0	0
	Técnico Básico					0	0	0	0	1	0
	Técnico Médio					0	0	0	0	0	1
	Nível Superior					0	0	0	0	0	0

Tabela A-2: Variáveis dummy (Continuação)

Variável	Categorias	Variável dummy								
		D _{pr1}	D _{pr2}	D _{pr3}	D _{pr4}	D _{pr5}	D _{pr6}	D _{pr7}	D _{pr8}	D _{pr9}

Sector Profissional	não indicado	1	0	0	0	0	0	0	0	0
	agricultura, silvicultura, pesca, extracção mineira	0	1	0	0	0	0	0	0	0
	indústria manufactureira	0	0	1	0	0	0	0	0	0
	transporte e comunicações	0	0	0	1	0	0	0	0	0
	sem ocupação	0	0	0	0	1	0	0	0	0
	comércio	0	0	0	0	0	1	0	0	0
	estudante	0	0	0	0	0	0	1	0	0
	educação, saúde	0	0	0	0	0	0	0	1	0
	administração	0	0	0	0	0	0	0	0	1
artesanato	0	0	0	0	0	0	0	0	0	

Tabela A-2: Variáveis dummy (Continuação)

Variável	Categorias	Variável dummy						
		D _{ag1}	D _{ag2}	D _{ag3}	D _{ag4}	D _{ag5}	D _{hist1}	D _{histl}
Agência	1	1	0	0	0	0		
	2	0	1	0	0	0		
	3	0	0	1	0	0		
	4	0	0	0	1	0		
	5	0	0	0	0	1		
	6	0	0	0	0	0		
Histórico com a instituição	Não possui crédito anterior						1	0
	Não possui atraso anterior (superior a 30dias)						0	1
	Possui atraso anterior (superior a 30 dias)						0	0