

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

CONSTRUÇÃO DE MODELOS DE PREVISÃO DE RISCO DE CRÉDITO

LISIANE PRISCILA ROLDÃO SELAU

Porto Alegre

2008

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

CONSTRUÇÃO DE MODELOS DE PREVISÃO DE RISCO DE CRÉDITO

Lisiane Priscila Roldão Selau

Orientador: Prof. Dr. José Luís Duarte Ribeiro

Banca Examinadora:

Prof. Dr. Cláudio José Müller
PPGEP / UFRGS

Prof. Dr. João Zani
PPGCC / UNISINOS

Profa. Dra. Márcia Elisa Soares Echeveste
PPGEP / UFRGS

Dissertação submetida ao Programa de Pós-Graduação em
Engenharia de Produção como requisito parcial à obtenção do título de
MESTRE EM ENGENHARIA DE PRODUÇÃO
Área de concentração: Sistemas de Qualidade

Porto Alegre

2008

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Professor Dr. José Luís Duarte Ribeiro
PPGEP / UFRGS
Orientador

Professor Dr. Flávio Sanson Fogliatto
Coordenador PPGEP / UFRGS

BANCA EXAMINADORA:

Professor Dr. Cláudio José Müller
PPGEP / UFRGS

Professor Dr. João Zani
PPGCC / UNISINOS

Professora Dra. Márcia Elisa Soares Echeveste
PPGEP / UFRGS

*“Mesmo que já tenhas feito uma longa
caminhada, há sempre um novo caminho a
fazer.”*

Santo Agostinho

DEDICATÓRIA

Dedico este trabalho ao meu esposo Jacques, com todo amor do meu coração, por sua paciência constante.

AGRADECIMENTOS

Ao professor Ribeiro pelo conhecimento transmitido e pela orientação e confiança neste trabalho.

Aos professores do PPGEP pelos aprendizados e em especial ao professor Fernando Amaral pelos conselhos de vida.

Ao professor Riboldi, meu eterno orientador.

Aos colegas de mestrado pelos momentos de descontração e trocas de experiências, em especial: Daniela, Joana, Mônica, Carlos, Diego, Giuliano e Marcelo, mais que colegas, amigos de verdade. Aos demais companheiros de trabalho do LOPP, em especial a Maria e ao Miorando, pelas dicas de formatação.

A Paula e ao Marcelo do LASTRAN que dedicaram muitas horas do seu tempo no auxílio com o *software* de redes neurais, vocês são ótimos.

Ao Mateus e a Adri, amigos para toda hora e para sempre.

Às amigas estatísticas Juliane, Dierê e Rosi; só nós para nos entendermos.

A minha irmã Lívian e ao meu cunhado Michel, pelo apoio e inspiração durante estes dois anos de estudo.

Aos funcionários da empresa em estudo, pela oportunidade e pelo tempo dispensado no desenvolvimento desta pesquisa.

Finalmente, agradeço em especial ao meu esposo amado, Jacques, que sempre me incentivou e apoiou em todas as decisões.

RESUMO

A presente dissertação tem como objetivo propor uma sistemática para a construção de modelos de previsão de risco de crédito e também comparar o desempenho de três técnicas estatísticas multivariadas utilizadas para sua construção: análise discriminante, regressão logística e redes neurais. O método proposto (denominado Modelo PRC) é composto de seis etapas: (i) delimitação da população; (ii) seleção da amostra; (iii) análise preliminar; (iv) construção do modelo; (v) escolha do modelo e (vi) passos para implantação. O Modelo PRC foi aplicado em uma amostra de 17.005 clientes de uma rede de farmácias com crediário próprio. Os resultados encontrados demonstram uma superioridade das redes neurais em relação às outras duas técnicas, o que era esperado devido a sua abordagem não-linear na combinação das variáveis. Considerando a venda anual aos clientes da base em estudo e utilizando o modelo neural desenvolvido, estima-se um acréscimo de 65% nos lucros.

Palavras-chave: Análise de Crédito – Modelo PRC – Análise Discriminante – Regressão Logística – Redes Neurais.

ABSTRACT

This work presents a methodology for credit risk prediction, comparing the performance of three statistical techniques used in the prediction process: discriminant analysis, logistic regression and neural networks. The proposed method (entitled PRC Model) embraces six steps: (i) population definition, (ii) sampling, (iii) preliminary analysis, (iv) model development, (v) model selection and (vi) implementation steps. The PRC Model was applied to a sample of 17,005 customers from an organization, which manages his own credit system and controls a pool of drugstores. The results show the superiority of neural networks over the other two techniques. This was expected due to the non-linear approach of the neural network when dealing with the explanatory variables. Considering the neural network model and the annual sales due to customers included on this study, the use of the proposed methodology indicates a 65% potential profits.

Keywords: Credit Analysis – PRC Model – Discriminat Analysis – Logistic Regression – Neural Networks

SUMÁRIO

LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1 INTRODUÇÃO	12
1.1 CONSIDERAÇÕES INICIAIS	12
1.2 TEMA E OBJETIVOS.....	14
1.3 JUSTIFICATIVA DO TRABALHO.....	15
1.4 METODOLOGIA DA PESQUISA	16
1.4.1 Caracterização da Pesquisa	16
1.4.2 Etapas do Trabalho.....	17
1.5 DELIMITAÇÃO DO TEMA.....	18
1.6 ESTRUTURA DO TRABALHO	19
2 REFERENCIAL TEÓRICO.....	21
2.1 CRÉDITO.....	21
2.1.1 Origens e Definições.....	21
2.1.2 Riscos de Crédito.....	23
2.1.3 Análise de Crédito.....	26
2.2 MODELOS DE PREVISÃO DE RISCO DE CRÉDITO	37
2.2.1 Análise Discriminante.....	41
2.2.2 Regressão Logística	46
2.2.3 Redes Neurais	49
2.2.4 Precisão dos Modelos na Predição.....	56
2.3 APLICAÇÕES DOS MODELOS DE PREVISÃO	60
2.4 CONCLUSÃO.....	65
3 SISTEMÁTICA PARA DESENVOLVIMENTO DO MODELO PRC.....	66
3.1 DELIMITAÇÃO DA POPULAÇÃO	66
3.2 SELEÇÃO DA AMOSTRA	68
3.3 ANÁLISE PRELIMINAR.....	70

3.4	CONSTRUÇÃO DO MODELO	71
3.5	ESCOLHA DO MODELO	73
3.6	PASSOS PARA IMPLANTAÇÃO	75
3.7	CONCLUSÃO	77
4	APLICAÇÃO DO MODELO PRC	78
4.1	DELIMITAÇÃO DA POPULAÇÃO	78
4.2	SELEÇÃO DA AMOSTRA	79
4.3	ANÁLISE PRELIMINAR	81
4.4	CONSTRUÇÃO DO MODELO PRC	84
4.4.1	Análise Discriminante	85
4.4.2	Regressão Logística	90
4.4.3	Redes Neurais	93
4.5	ESCOLHA DO MODELO	97
4.6	PASSOS PARA IMPLANTAÇÃO	101
4.7	AVALIAÇÃO GERENCIAL	106
4.8	AVALIAÇÃO DE GANHO POTENCIAL	108
4.9	DISCUSSÃO DOS RESULTADOS	109
5	CONSIDERAÇÕES FINAIS	112
5.1	CONCLUSÕES	112
5.2	SUGESTÕES PARA TRABALHOS FUTUROS	114
	REFERÊNCIAS	115
	APÊNDICE A - AGRUPAMENTO DE PROFISSÕES	121
	APÊNDICE B - AGRUPAMENTO DE CIDADES DE NASCIMENTO	122
	APÊNDICE C - AGRUPAMENTO DE CEP RESIDENCIAL	123
	APÊNDICE D - AGRUPAMENTO DE CEP COMERCIAL	124
	APÊNDICE E – PESOS DOS NEURÔNIOS DA REDE NEURAL	125
	APÊNDICE F – RELATÓRIO DE ACOMPANHAMENTO DE PERFIL	126
	ANEXO A – FORMULAÇÃO DA RÉPLICA DA REDE NEURAL EM PLANILHA ELETRÔNICA	128

LISTA DE FIGURAS

Figura 1	Representação do crédito	22
Figura 2	Taxas de crescimento do crédito (variação em doze meses)	30
Figura 3	Contraste entre crédito a pessoas físicas e pessoas jurídicas	31
Figura 4	Quadro resumo dos Cs do crédito	32
Figura 5	Estágios de decisão da análise discriminante	43
Figura 6	Representação de escores Z discriminantes	45
Figura 7	Forma da relação logística entre as variáveis	47
Figura 8	Modelo não-linear de um neurônio	51
Figura 9	Modelo de uma rede neural	51
Figura 10	Treinamento versus capacidade de generalização da rede neural	54
Figura 11	Exemplificação de um ponto de corte	57
Figura 12	Análise objetiva e subjetiva aliadas na concessão de crédito	58
Figura 13	Ponto de corte com tamanhos diferentes de amostras	59
Figura 14	Etapas para desenvolvimento do Modelo PRC	66
Figura 15	Verificação de acerto nas classificações do modelo	74
Figura 16	Exemplificação do método de obtenção do valor do teste KS	75
Figura 17	Exemplo de política de crédito segundo a inadimplência	76
Figura 18	Variáveis identificadas para criação do modelo	79
Figura 19	Taxa de sinistro e separação do modelo discriminante	87
Figura 20	Distribuição dos escores da função discriminante	89
Figura 21	Taxa de sinistro e separação do modelo logístico	92
Figura 22	Taxa de sinistro e separação do modelo neural	95
Figura 23	Representação do valor de KS para os modelos construídos	101
Figura 24	Fluxograma para análise de crédito usando o Modelo PRC	105

LISTA DE TABELAS

Tabela 1	Comparativo de indicadores de crédito e renda no Brasil.....	12
Tabela 2	Classificação de risco.....	26
Tabela 3	Total de clientes por tipo	80
Tabela 4	Risco relativo de variáveis excluídas da análise	81
Tabela 5	Risco relativo da variável salário.....	81
Tabela 6	Exemplo de agrupamento de atributos segundo risco relativo.....	82
Tabela 7	Agrupamento final e variáveis <i>dummies</i> segundo risco relativo.....	83
Tabela 8	Criação de variáveis <i>dummies</i> para demais variáveis independentes	84
Tabela 9	Coeficientes da função discriminante canônica	85
Tabela 10	Distribuição de clientes e taxa de sinistro do modelo discriminante....	86
Tabela 11	Comparação entre amostras segundo modelo discriminante	88
Tabela 12	Resultado do teste Box's M.....	89
Tabela 13	Coeficientes da regressão logística	90
Tabela 14	Distribuição de clientes e taxa de sinistro do modelo logístico.....	91
Tabela 15	Comparação entre amostras segundo modelo logístico	93
Tabela 16	Comparação dos melhores modelos neurais construídos.....	94
Tabela 17	Distribuição de clientes e taxa de sinistro do modelo neural.....	95
Tabela 18	Comparação entre amostras segundo modelo neural	96
Tabela 19	Percentuais de acerto do modelo discriminante.....	97
Tabela 20	Percentuais de acerto do modelo logístico.....	98
Tabela 21	Percentuais de acerto do modelo neural.....	99
Tabela 22	Classificação dos grupos segundo os modelos construídos.....	99
Tabela 23	Valores de KS para os modelos construídos	100
Tabela 24	Sensibilidade do ponto de corte do modelo discriminante	102
Tabela 25	Sensibilidade do ponto de corte do modelo logístico	103
Tabela 26	Sensibilidade do ponto de corte do modelo neural	103
Tabela 27	Média de compras dos clientes no período.....	108
Tabela 28	Estimativas de venda e ganho para clientes da amostra	109

1 INTRODUÇÃO

Nesta seção são apresentadas as considerações iniciais, os objetivos, a justificativa do trabalho, sua metodologia e estrutura.

1.1 CONSIDERAÇÕES INICIAIS

O crédito no Brasil sempre foi escasso e, devido a políticas mal concebidas e ao processo inflacionário do passado, a cultura do crédito como instrumento para crescimento dos negócios ainda está em estágio inicial. Entretanto, em consequência da maior estabilidade da economia brasileira nos últimos anos, as empresas têm percebido o crédito como um gerador de riquezas e de novos negócios (GOLDBERG *apud* BUENO, 2003).

Segundo Pereira (2006), o crescimento vertiginoso do crédito nos últimos anos vem alertando os analistas do setor sobre um período de turbulência que está por vir. A Tabela 1 apresenta informações sobre o crescimento do crédito e da renda mensal dos brasileiros. Observa-se uma disparidade entre os números, sendo o aumento na demanda por crédito maior que o aumento na renda média.

Tabela 1
Comparativo de indicadores de crédito e renda no Brasil

	Out./04	Out./05	Crescimento
Crédito Pessoa Física (R\$ MM)	108,4	151,5	39,8%
Renda média mensal (R\$)	949,24	966,1	1,8%

Fonte: Pereira (2006)

O crescimento da demanda por crédito à pessoa física no Brasil vem revolucionando esse mercado, fazendo com que as empresas do setor se moldem para ficarem à altura das oportunidades. O fenômeno provocou uma reengenharia nos sistemas de crédito em relação à tecnologia na concessão (PEREIRA, 2006).

As empresas que concedem crédito estão apostando numa melhor análise de crédito, evitando trabalhar com clientes que ofereçam risco, diminuindo o índice de inadimplência. Por isso estão utilizando, além da experiência do analista, métodos e técnicas que auxiliam na tarefa de decidir se um cliente é merecedor de crédito. A gestão de risco passou a ocupar, nos últimos tempos, posição de destaque na administração financeira, especialmente em consequência da expansão do crédito, do crescimento do mercado e da globalização (BUENO, 2003).

Segundo Steiner *et al.* (1999), a correta decisão de crédito é essencial para a sobrevivência das instituições financeiras. Qualquer erro na decisão de conceder o crédito pode significar que, em uma única operação, haja a perda do ganho obtido em dezenas de outras bem sucedidas, já que o não recebimento representa a perda total do montante emprestado. Portanto, é importante prever e reduzir a inadimplência, pois os prejuízos com créditos mal sucedidos deverão ser cobertos com a cobrança de altas taxas de juros em novas concessões.

Conforme afirma Schrickel (1997), a análise de risco envolve a habilidade de fazer uma decisão de crédito, dentro de um cenário de incertezas e constantes mutações e informações incompletas. Esta habilidade depende da capacidade de analisar logicamente situações, muitas vezes complexas, e chegar a uma conclusão prática e factível de ser implementada.

Em muitas empresas, a avaliação de crédito é feita com base em uma enorme variedade de informações vindas das mais diversas fontes. Os gerentes analisam estas informações de maneira subjetiva e muitas vezes não conseguem explicar os processos de tomada de decisões, embora consigam apontar os fatores que influenciam as decisões. Além disso, estes ambientes são dinâmicos, com constantes alterações, onde as decisões devem ser tomadas rapidamente (MENDES FILHO *et al.*, 1996).

Em alguns casos é vital o uso dos modelos de previsão de risco, que baseados em dados recentes de clientes com a empresa, geram uma pontuação para as características que levam a criação de um padrão de comportamento em relação à inadimplência. Segundo Guimarães e Chaves Neto (2002), quando a empresa tem à sua disposição uma regra de reconhecimento de padrões e

classificação que indique previamente a chance de inadimplência de um futuro cliente, a decisão de concessão de crédito fica muito facilitada, podendo-se então utilizar argumentos quantitativos em substituição aos argumentos subjetivos e decidir com maior confiança.

1.2 TEMA E OBJETIVOS

Este trabalho aborda três temas principais: concessão de crédito, risco e modelos de previsão. A concessão de crédito é um importante instrumento para o desenvolvimento econômico e também se constitui como a principal atividade de instituições financeiras. O risco de crédito ou inadimplência pode ser definido como a possibilidade do não cumprimento das obrigações contratuais relativas às transações financeiras. Os modelos de previsão têm a finalidade de estimar este risco com base nos dados cadastrais do cliente, utilizando um sistema de pontuação com uso de técnicas estatísticas multivariadas para identificação de padrões de comportamento quanto à inadimplência (SILVA, J., 2003).

O objetivo deste trabalho é propor uma sistemática para a construção de modelos de previsão de risco de crédito, também denominados *Credit Scoring*, para concessão de crédito a pessoas físicas. Além do objetivo geral, são estabelecidos os seguintes objetivos específicos:

- comparar desempenho, vantagens e desvantagens de três técnicas estatísticas multivariadas utilizadas para a construção do modelo (análise discriminante, regressão logística e redes neurais);
- analisar a percepção da gerência de crédito quanto à implementação do modelo de previsão no sistema da empresa;
- avaliar o potencial aumento dos lucros da empresa com a utilização do modelo de previsão de risco de crédito.

1.3 JUSTIFICATIVA DO TRABALHO

O tema se encaixa no contexto das decisões financeiras de qualquer tipo de empresa (industriais, comerciais ou de serviços), já que a concessão de crédito é uma forma de empréstimo ou de financiamento que muitas fornecem aos seus clientes. Segundo Silva, M. (2003), a arte de bem conceder crédito é fundamental tanto para quem concede como para quem recebe. No primeiro caso, o retorno dos recursos emprestados é fator determinante para novas concessões e, muitas vezes, a sobrevivência do próprio negócio. No segundo, poderá ser a solução para avançar nos negócios ou mesmo resolver uma situação financeira desfavorável.

Segundo Securato (1996), na área financeira das indústrias, do comércio e de demais atividades econômicas, os homens de finanças estão constantemente envolvidos com tomadas de decisão que podem representar o sucesso ou fracasso próprio, de um projeto específico ou mesmo da empresa. Conceder ou não crédito, tomar recursos ou aplicá-los são decisões constantes a que estão sujeitos os executivos financeiros, qualquer que seja o negócio da empresa.

Na área financeira, o risco está presente em quase totalidade das decisões. Conhecer e prever este risco auxilia o processo decisório, facilitando a tomada de decisão. Leme *apud* Securato (1996, p. 17) define previsão como o “processo pelo qual a partir de informações existentes, admitidas certas hipóteses e através de algum método de geração, chegamos a informações sobre o futuro, com uma determinada finalidade.

Saber se um cliente provavelmente honrará com seus compromissos apresenta-se como uma informação imprescindível na hora de tomar uma decisão com vistas à concessão de crédito. Com isso, pode-se demonstrar que as instituições financeiras poderiam ter um acréscimo nos lucros se, na concessão de crédito, os critérios fossem mais rígidos. De posse da classificação fornecida pelo modelo, a empresa tem um diagnóstico preliminar do provável comportamento de um novo cliente, negando ou não a concessão do crédito (VASCONCELLOS, 2002).

Os modelos de previsão quantitativos são muito utilizados para auxílio na análise de crédito, tendo como vantagens o aumento do número de mercedores

que terão o crédito aprovado, aumentando os lucros; o aumento do número de não merecedores que não receberão o crédito, diminuindo as perdas; as solicitações de crédito podem ser analisadas rapidamente; os critérios subjetivos são substituídos pelas decisões objetivas; e por fim um menor número de pessoas será necessário para administrar o crédito (ROSENBERG; GLEIT, 1994; STEINER *et al.*, 1999). Segundo Hand e Henley (1997), se a concessão do crédito for muito lenta é possível que clientes que não necessitem com urgência ou tenham outras opções de crédito se cansem e desistam da solicitação.

Para Silva, J. (2003), as técnicas estatísticas multivariadas têm sido consideradas como ferramentas bastante poderosas na administração do risco de inadimplência existente na concessão de crédito. Para construção de um modelo de previsão de risco de crédito é importante identificar qual a técnica estatística mais eficiente para modelar os dados da empresa de forma a conseguir a melhor previsão de comportamento dos clientes.

1.4 METODOLOGIA DA PESQUISA

1.4.1 Caracterização da Pesquisa

Pesquisa pode ser definida como um conjunto de ações propostas para encontrar a solução para um problema, tendo por base procedimentos racionais e sistemáticos. Realiza-se pesquisa quando se tem um problema e não se têm informações para solucioná-lo. Os tipos de pesquisa podem ser classificados em função de sua natureza, de sua forma de abordagem, de seus objetivos e de seus procedimentos técnicos (SILVA; MENEZES, 2005).

O método de pesquisa utilizado neste trabalho, no que diz respeito a sua natureza, trata-se de pesquisa aplicada, pois objetiva buscar conhecimentos para aplicação prática com vistas à solução de problemas específicos.

Quanto à abordagem do problema, esta dissertação pode ser vista como uma pesquisa essencialmente quantitativa, já que se propõe estudar modelos para previsão de risco de crédito e redução de inadimplência com uso de técnicas

estatísticas. No entanto, ao final do estudo, é apresentada uma avaliação qualitativa da gerência da empresa quanto à aceitação do uso da tecnologia.

Segundo os objetivos, a presente dissertação pode ser classificada com uma pesquisa explicativa, pois visa descobrir os fatores (características dos clientes) que determinam ou contribuem para a ocorrência dos fenômenos (inadimplência).

Do ponto de vista dos procedimentos, a pesquisa apresentada nesta dissertação é desenvolvida através de uma pesquisa-ação por ser realizada com o envolvimento da empresa, de forma a buscar a resolução de um problema coletivo (conceder crédito com menor risco). O método de pesquisa empregado compreende os seguintes pontos: (i) revisão e entendimento da teoria; (ii) desenvolvimento de um modelo para previsão do risco de crédito; (iii) aplicação do modelo desenvolvido; (iv) avaliação do modelo proposto.

1.4.2 Etapas do Trabalho

O presente trabalho é desenvolvido através das etapas descritas a seguir:

- a. buscar referencial teórico sobre a importância e utilização da análise de crédito como medida para controle de risco de inadimplência, bem como o uso de modelos de previsão de risco de crédito baseados em padrões de comportamento. Entender técnicas estatísticas utilizadas para construção destes modelos (análise discriminante, regressão logística e redes neurais);
- b. contatar empresa para entender o cenário e as variáveis envolvidas com relação ao que a literatura recomenda. O entendimento é essencial para a construção da metodologia a ser utilizada;
- c. detalhar passos para o estudo do modelo de previsão de risco de crédito, desde a coleta de dados até a conclusão do modelo para implantação no sistema da empresa;
- d. definir para a empresa parâmetros para a extração da amostra: tamanho, período e variáveis. Validar consistência e preenchimento dos dados

(*Data Cleaning*), fazendo descarte de observações atípicas. Separar parte da amostra para teste dos modelos;

- e. fazer análises preliminares de cada variável para criação de variáveis *dummies*. Aplicar as três técnicas em estudo e verificar se os resultados obtidos são consistentes;
- f. verificar o desempenho dos modelos através da amostra de teste, que não é utilizada no momento do desenvolvimento dos modelos. Esta amostra de teste estaria, desta forma, simulando a situação de análise e concessão de crédito a novos clientes;
- g. avaliar o potencial aumento nos lucros da empresa, levando em consideração o fato de que alguns créditos não teriam sido concedidos aos clientes que o modelo sinalizar como potenciais inadimplentes;
- h. apresentar resultados obtidos à empresa e avaliar percepção quanto à utilização dos modelos de previsão de risco de crédito em comparação com a análise de crédito baseada em critérios subjetivos.

1.5 DELIMITAÇÃO DO TEMA

De acordo com Rosenberg e Gleit (1994), os modelos de previsão de risco de crédito estão em duas categorias: modelos de aprovação de crédito (*Credit Scoring*) e modelos de escoreamento comportamental (*Behavioural Scoring*). Segundo Thomas (2000) a principal diferença entre os modelos é que o primeiro é utilizado como suporte à decisão de concessão de crédito a novos clientes, enquanto que o segundo serve de auxílio para a administração de contas de clientes que já tem relação de crédito com a empresa. Neste trabalho trata-se do estudo de modelos para aprovação de crédito a novos clientes de uma empresa que concede crédito próprio para financiamento dos produtos vendidos em suas lojas.

O foco deste trabalho está nos modelos de previsão de risco de crédito a pessoas físicas na modalidade de financiamento do setor varejista para a localidade onde está estabelecida a empresa em estudo. Os modelos resultantes das análises

são específicos para o caso estudado, não devendo ser estendidos a outras empresas com características diversificadas (formas de parcelamento, perfil de clientes, bens adquiridos, etc.). Neste caso, faz-se necessário construir um modelo de previsão avaliando o perfil de comportamento específico da carteira de clientes.

O modelo com melhor desempenho não chegou a ser implantado na empresa até o final deste trabalho. Portanto, os ganhos potenciais com a utilização do modelo na concessão do crédito somente foram mensurados através de amostras de testes.

Uma limitação existente em qualquer formulação de modelos de pontuação está relacionada à obtenção de uma amostra com viés de seleção, já que a população estudada refere-se somente a créditos concedidos. Os clientes que foram negados por algum motivo e que, portanto, não farão parte da amostra, serão potenciais clientes para solicitações de crédito futuras, mas seu perfil pode não estar contemplado no modelo.

Vasconcellos (2002) apresenta um estudo sobre o efeito desse viés em modelos de previsão, mostrando que o uso de amostras restritas aos créditos aprovados gera resultados com vieses, mas que o tamanho e direção do viés do modelo nem mesmo podem ser conhecidos. O autor alerta também que não seria interessante uma empresa incorrer nos custos de conceder crédito a todos os clientes, pois ainda assim não se chegaria à população de todos os potenciais tomadores de crédito, já que poderia ocorrer um viés por parte dos próprios clientes que podem escolher a empresa para solicitar crédito de acordo com seus critérios. Feelders (2000), em seu artigo, também relata estudos para tentar inferir o comportamento dos proponentes rejeitados pela instituição e que, portanto, não fazem parte da amostra. O método utilizado pelo autor, denominado por inferência dos rejeitados, consiste em inferir o comportamento dos proponentes rejeitados, casos tivessem sido aprovados.

1.6 ESTRUTURA DO TRABALHO

Com vistas a atingir os objetivos propostos, esta dissertação está dividida em cinco capítulos, conforme apresentado a seguir.

No primeiro capítulo é apresentada a introdução com a contextualização dos temas crédito, risco e modelos de previsão, abordando os objetivos que se buscam alcançar, bem como suas justificativas. Também se limita o alcance da pesquisa, de forma a esclarecer os pontos que não são abordados, apesar de importantes.

O segundo capítulo trata do referencial teórico, onde são abordados temas associados à importância e utilização da análise de crédito como medida para controle de risco de inadimplência, bem como o uso de métodos para previsão de risco baseados em padrões de comportamento. São apresentadas e confrontadas as técnicas estatísticas utilizadas para construção destes modelos de pontuação: análise discriminante, regressão logística e redes neurais.

A metodologia utilizada para construção de um modelo de pontuação para concessão de crédito, desde a escolha da população até os passos para a implementação no sistema da empresa, é apresentada no terceiro capítulo.

No quarto capítulo são apresentados os resultados da construção e aplicação dos modelos e da validação técnica (realizada através da comparação das três técnicas estatísticas utilizadas). Também se avalia a aceitação da gerência, ou seja, sua percepção referente ao uso do modelo em comparação com a concessão de crédito subjetiva. Por fim, são apresentados os resultados de simulação de utilização do modelo de previsão, de forma a verificar o potencial aumento nos lucros da empresa.

O quinto capítulo traz as principais conclusões obtidas e a verificação do alcance dos objetivos, bem como a justificativa do assunto. Também são sugeridas possibilidades de trabalhos futuros.

2 REFERENCIAL TEÓRICO

2.1 CRÉDITO

Crédito é um conceito presente no dia-a-dia das pessoas e empresas mais do que possamos imaginar a princípio. Todos nós, tanto as pessoas quanto as empresas, estamos continuamente às voltas com o dilema de uma equação simples: a constante combinação de nossos recursos finitos com o conjunto de nossas imaginações e necessidades infinitas – existem mais maneiras de se gastar dinheiro, por exemplo, do que de ganhá-lo (SCHRICKEL, 1997, p. 11).

2.1.1 Origens e Definições

Crédito tem origem no latim *credere*, que significa crer, confiar, acreditar, ou ainda do substantivo *creditum*, que literalmente significa confiança (BLATT, 1998). A palavra crédito pode ter vários significados, dependendo do contexto do qual se esteja tratando, mas em finanças, de acordo com Silva, J. (2003), crédito é um instrumento de política financeira utilizado por empresas comerciais ou industriais na venda a prazo de seus produtos, ou por um banco comercial na concessão de empréstimo ou financiamento.

Define-se crédito como o ato de vontade ou disposição de alguém ceder, temporariamente, parte do seu patrimônio a um terceiro, com a expectativa de que esta parcela volte a sua posse integralmente depois de decorrido o tempo previamente estipulado. Esta parte do patrimônio pode ser materializada por dinheiro (empréstimo monetário) ou bens (venda com pagamento parcelado ou a prazo). Sendo um ato de vontade, sempre caberá ao cedente do patrimônio a decisão de cedê-lo ou não (SCHRICKEL, 1997). Tal relação entre as partes é ilustrada na Figura 1, identificando o significado restrito do crédito.

Conforme Silva (1988), essa função de intermediário financeiro coloca o crédito como o fator de maior importância numa instituição financeira, de forma que as operações de crédito constituem a essência do seu próprio negócio. A concessão de crédito, portanto, consiste em colocar à disposição do cliente determinado valor

monetário em determinado momento, mediante promessa de pagamento futuro, tendo como retribuição por essa prestação de serviço determinada taxa de juros.

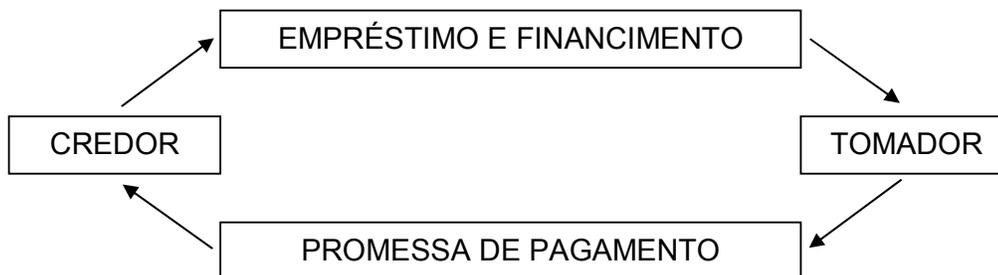


Figura 1 Representação do crédito

Fonte: adaptado de Silva, J. (2003, p. 63)

Em grande parte do mundo, a concessão de crédito é baseada em facilidades. Os consumidores esperam ser capazes de comprar a maioria de seus produtos e serviços a crédito. Para as empresas, a concessão de crédito representa uma ferramenta essencial para efetuar vendas. As atividades de marketing e vendas enfatizam como é simples comprar a crédito. Dessa forma, empresas acabam sendo forçadas a efetuarem a maioria de suas vendas a crédito (BLATT, 1998).

No comércio, de um modo geral, o crédito assume o papel de facilitador da venda, possibilitando ao cliente adquirir o bem para atender sua necessidade, ao mesmo tempo em que incrementa as vendas do comerciante. Na indústria, o crédito também assume o papel de facilitador de venda. Um fabricante de equipamentos hospitalares, por exemplo, pode abrir linhas de crédito para venda de seus produtos e, com isso, possibilitar a vários médicos, clínicas e hospitais a aquisição de seus equipamentos. Sem a alternativa do crédito, a quantidade de compradores poderia ser muito menor e, por consequência, o lucro do fabricante também seria reduzido (SILVA, J. 2003).

Segundo Almeida e Siqueira (1997), as empresas têm buscado ferramentas e mecanismos que possam agregar algum diferencial mercadológico e proporcionar uma maior rentabilidade e segurança nas suas relações comerciais. Com a estabilização da moeda, os bancos brasileiros foram obrigados a passar por um

processo rigoroso de adaptação. A forte e rápida redução dos ganhos inflacionários forçou os bancos a aumentarem seus volumes de crédito.

Silva, J. (2003) afirma que o crédito tem um papel econômico e social muito importante na vida das empresas e das pessoas. Possibilita às empresas aumentarem seu nível de atividade, estimula o consumo influenciando na demanda, ajuda as pessoas a obterem moradia, bens e até alimentos; e facilita a execução de projetos para os quais as empresas não disponham de recursos próprios suficientes. Por outro lado, cabe salientar que o uso inadequado do crédito pode levar uma empresa à falência ou um indivíduo à insolvência, assim como pode ser parte componente de um processo inflacionário.

É normal associar às transações de empréstimo um preço remuneratório, que deve ser pago pelo tomador ao credor. A este preço dá-se o nome de juros ou preço do dinheiro, que se presta à compensação dos riscos assumidos pelo credor quanto à possível perda ou deterioração da parcela de seu patrimônio cedido. Quando esse preço é expresso em forma percentual, tem-se a taxa de juros, também conhecida como taxa de risco (SCHRICKEL, 1997).

2.1.2 Riscos de Crédito

Segundo Silva (1988), para as instituições de crédito em geral, cujo principal produto é o dinheiro, não há possibilidade de venda à vista. Isto torna o crédito e a avaliação de risco do negócio condição essencial de sobrevivência e realização de resultados positivos.

Toda vez que uma instituição financeira vende um crédito está automaticamente comprando um risco com todos os efeitos bons e ruins que a transação envolve. Diante dessa relação entre crédito e risco, alguns autores como Caouette *et al.* (1999) e Santos (2000) chegam a explicitar a existência do risco de crédito em suas definições de crédito, ao definir crédito como a troca de um valor presente por uma promessa de reembolso futuro, não necessariamente certo, em virtude do fator risco.

Apesar do risco de crédito ser uma das formas mais antigas de risco nos mercados financeiros, executivos e acadêmicos ainda debatem, acaloradamente, a melhor forma de sua apuração e administração. A gestão do risco de crédito continua a ser substancialmente uma atividade quase caseira, em que as decisões quanto a empréstimos são feitas sob medida, caso a caso (CAOJETTE *et al.*, 1999).

Para Schrickel (1997), risco significa incerteza, imponderável, imprevisível, e estes se situam necessária e unicamente no futuro. Assim, embora a análise de crédito deva lidar com eventos passados do tomador de empréstimos (a análise histórica), as decisões de crédito devem considerar primordialmente o futuro desse tomador. O risco situa-se no futuro; no passado, encontra-se apenas história.

O risco é constituído pela ocorrência de qualquer fato adverso para uma dada situação esperada. Em específico, o risco de crédito pode ser definido como a probabilidade de ocorrência de perdas por inadimplência com relação a situações desejáveis, no caso o retorno integral dos créditos por parte da totalidade dos devedores (BORGES; BERGAMINI, 2001).

De acordo com Silva, J. (2003), alguns autores utilizam as palavras risco e incerteza indistintamente, como se tivessem significados idênticos. Para Borges e Bergamini (2001), risco deve ser diferenciado de incerteza, embora ambos os termos definam possíveis eventos que possam ocorrer no futuro. O primeiro refere-se a fatos recorrentes ou repetitivos, permitindo sua medição com base em uma distribuição de probabilidades e, em decorrência, a sua estimação com razoável grau de precisão. Já a incerteza tem uma incidência aleatória por envolver fatos não recorrentes, impedindo a mensuração da probabilidade de sua ocorrência, ou pelo menos a sua estimação com utilidade para tomada de decisões referentes ao crédito.

O risco de crédito está diretamente ligado ao mercado e suas mudanças. A gestão deve acompanhar essas flutuações para que a cultura do crédito e as estratégias de empréstimos possam ser repensadas e até redesenhadas (CAOJETTE *et al.*, 1999). Segundo Schrickel (1997), o risco sempre estará presente em qualquer empréstimo. Não há empréstimo sem risco, porém o risco deve ser compatível com o negócio da empresa e à sua margem mínima almejada (receita).

Borges e Bergamini (2001) apontam outro ponto importante na administração do risco de crédito, a proteção jurídica ao devedor. Segundo os autores, existe um ambiente processual de proteção excessiva ao devedor. Isto faz com que a liquidação financeira resulte num baixo nível de recuperação, o que é resultado tanto das características legais dos instrumentos utilizados quanto da morosidade do aparato jurídico em resolver conflitos. Isto justifica o dito corrente no mercado financeiro nacional de que 'o credor é refém do devedor'. Neste sentido, verifica-se que no momento que um devedor resolver não pagar e questionar a dívida judicialmente estará amparado pela legislação brasileira.

Borges e Bergamini (2001) definem inadimplência como a característica de um devedor que não pagou uma obrigação na data fixada ou mesmo que ingressou em curso problemático. Por crédito problemático entende-se a ocorrência do não pagamento por prazos maiores vinculados a crises de liquidez, sendo definido pela empresas como perda ou crédito irrecuperável. Para Eifert (2003), o inadimplemento caracteriza-se pela falta ou atraso no cumprimento de uma obrigação ou prestação vencida. Assim, o descumprimento dos pagamentos contratados já estaria contribuindo para o aumento do risco do crédito.

De acordo com Santos (2000), o risco de inadimplência pode ser determinado por fatores internos e externos. Fatores internos de risco são aqueles voltados à falta de experiência dos administradores no gerenciamento do crédito, a controles inadequados, à concentração de crédito nas mãos de clientes de alto risco, à falta de modelagem estatística e à política estratégica de crédito da instituição. São fatores, segundo o autor, controláveis, todavia, dependentes do nível de formação, da experiência adquirida e da especialização técnica dos empresários.

Os fatores externos de risco são de natureza macroeconômica, e segundo Santos (2000), são os eventos que afetam o sistema econômico da empresa, não sendo controláveis por ela. Acontecimentos como ações do governo, desemprego, inflação, conjuntura econômica e recessão são alguns exemplos de fatores externos que podem influenciar a capacidade de pagamento.

Uma classificação de riscos de crédito constitui uma ferramenta importante para obter uma melhor distinção entre as possíveis situações relacionadas à

concessão do crédito. Na Tabela 2, é definida a classificação de riscos de crédito apresentada por Silva, J. (2003), baseada na Resolução 2.682 do Banco Central com uma escala que vai de AA até H. Essa classificação foi elaborada com base na qualidade dos devedores, tipo, destinação e valor da operação. O intuito desta classificação é funcionar como uma central de risco, auxiliando a decisão do crédito. A classe AA é a que representa o menor risco, para a qual não há provisão de inadimplência, sem atraso nos pagamentos. Já na classe H, tem-se o extremo da classificação, onde 100% das operações devem ser provisionadas.

Tabela 2
Classificação de risco

Dias de Atraso	Nível de Risco	% de Provisão
0	AA	0,0
1 a 14	A	0,5
15 a 30	B	1,0
31 a 60	C	3,0
61 a 90	D	10
91 a 120	E	30
121 a 150	F	50
151 a 180	G	70
181 ou +	H	100

Fonte: Adaptado de Silva, J. (2003, p. 76 e 77)

Para Silva, J. (2003), a classificação de risco é uma das várias ferramentas que podem ser utilizadas pelas áreas de análise de crédito para a tomada de decisões. De acordo com Santos (2000), a implementação de escalas é que possibilita aos credores adotar preços e cobrar juros compatíveis com a faixa de risco de cada cliente.

Saunders (2000) constata que as metodologias utilizadas para análise de risco ainda estão em fase de aprimoramento. Existem muitas lacunas na busca de uma gestão adequada de risco de crédito, em meio a uma geração de profissionais de engenharia financeira que vêm aplicando seus conhecimentos e habilidades em construção de modelos para análise desta área.

2.1.3 Análise de Crédito

Numa instituição financeira (como para qualquer prestador), o objetivo da análise de crédito é identificar os riscos nas operações de empréstimo. Para a

identificação dos riscos é necessário evidenciar conclusões quanto à capacidade de pagamento do tomador e fazer recomendações relativas à melhor estruturação e tipo de empréstimo a conceder, à luz das necessidades financeiras do solicitante, dos riscos identificados, sob a perspectiva de maximização dos resultados da instituição (SCHRICKEL, 1997).

Ainda de acordo com o mesmo autor, os instrumentos de análise variam com a situação peculiar que se tem à frente. Porém, tomar uma decisão dentro de um contexto incerto, em constante mutação, e tendo em mãos um volume de informações nem sempre suficiente, é extremamente difícil. Portanto, esta decisão será tanto melhor, quanto melhores forem as informações disponíveis.

Segundo Schrickel (1997), a atividade de concessão de crédito, como tantas outras, baseia-se em informações e decisões. A esse respeito, Silva J. (2003) destaca que a obtenção de informações confiáveis e seu tratamento constituem a base para uma decisão de crédito segura. Antes da tomada de decisão, cabe ao credor realizar o processo de análise de crédito, buscando o maior número de informações relevantes, juntamente com o tratamento adequado destes dados.

Em princípio, qualquer crédito só deve ser concedido sob amparo de uma aprovação específica, dada em consequência do cumprimento da metodologia de aprovação de crédito adotada pela instituição. As condições gerais e específicas de cada operação de crédito deverão sempre espelhar as negociações empreendidas entre o credor e o cliente, devendo refletir a realidade do mercado à época de sua efetiva negociação e concessão (SCHRICKEL, 1997).

Neste sentido, Lewis (1992) utiliza o conceito de *Creditworthiness* para indicar a viabilidade de um cliente ou de uma operação. Afirmar que um cliente ou operação é digno de crédito significa que são financeiramente interessantes e, portanto, devem ser aprovados em um processo de concessão de crédito. Vasconcellos (2004) destaca que a missão de um processo de avaliação de crédito é identificar as operações financeiramente viáveis, levando em consideração as características do cliente, sua situação atual e os detalhes da operação.

Segundo Silva, J. (2003), alguns bancos copiam os formulários de análise de outros bancos e passam a usar os mesmos indicadores e a mesma forma de análise para uma carteira de clientes com perfil diferente, levando ao uso de padrões e critérios inadequados. A eficiência da instituição financeira na concessão do crédito, através de critérios de seleção de clientes e de avaliação dos respectivos riscos, é que determinará seus resultados, afetando sua solidez e rentabilidade.

Schricket (1997) afirma que perder dinheiro faz parte do negócio de crédito, mas o que jamais deve ocorrer é que a perda tenha ocorrido por informações que não foram devidamente ponderadas, embora previstas ou previsíveis. A esta perda o autor dá o nome de 'perda mal perdida' ou 'perda burra'.

2.1.3.1 Análise de Crédito Subjetiva x Objetiva

O processo de análise de crédito pode recorrer a técnicas subjetivas ou a técnicas objetivas. A análise de crédito subjetiva, também conhecida como análise tradicional de crédito, envolve decisões individuais quanto à concessão ou recusa de crédito. Esta decisão baseia-se na disponibilidade de informações, no conhecimento especializado e sensibilidade de cada analista para a aprovação do crédito. As técnicas objetivas fazem uso da estatística para obtenção de modelos para previsão do perfil dos clientes (SANTOS, 2000).

As regras que são utilizadas para fazer a análise de crédito são comumente criadas internamente, sendo, portanto, uma função da cultura de crédito de uma instituição (CAOQUETTE, 1999). Para Silva (1988), políticas são instrumentos utilizados pelas empresas com o objetivo de determinar padrões de decisão na resolução de problemas semelhantes. Tais políticas devem ser estabelecidas observando-se os objetivos estabelecidos pela empresa, as regras governamentais e a capacidade de aplicação e captação de recursos.

Durante muito tempo, a capacidade de julgamento de um analista foi considerada uma importante característica de um administrador. Porém, atualmente, isto não é suficiente, haja vista a evolução da ciência e dos métodos científicos que

vem sendo transmitidos com o passar dos anos. Tal aperfeiçoamento progressivo constitui-se numa ferramenta muito útil na tomada de decisão (SILVA, J. 2003).

Saunders (2000) afirma que é difícil distinguir entre abordagens tradicionais e abordagens novas (modelos de gestão de carteiras) de análise de crédito, dado que muitas das melhores idéias dos modelos tradicionais são utilizadas nos modelos novos. O autor sugere três classes de modelos, compreendendo a abordagem tradicional:

- sistemas especialistas – a decisão de crédito envolve fatores objetivos e subjetivos, ficando a cargo do analista de crédito que tem em sua experiência um poderoso instrumento. O conhecimento especializado, seu julgamento subjetivo, e sua atribuição de pesos a certas variáveis são as mais importantes determinantes na decisão de conceder ou não o crédito. Um dos sistemas especialistas mais comuns são os Cs do crédito;

- sistemas de classificação ou *rating* – separam os empréstimos em diferentes classes de acordo com a probabilidade de perda. Segundo esta classificação, são provisionados valores para perdas, de forma que quanto melhor for o *rating*, menor será o percentual de provisionamento do empréstimo. Um exemplo de sistema de classificação é o sugerido pelo Banco Central na Resolução 2.682, que foi apresentado na Tabela 2;

- sistemas de pontuação de crédito – identificam fatores que determinam a probabilidade de inadimplência, e sua combinação ou ponderação para produzir uma pontuação quantitativa. Podem ainda ser divididos entre modelos para concessão de crédito (*Credit Scoring*) e modelos para comportamento de clientes já existentes (*Behavioural Scoring*).

A análise clássica de crédito é um sistema especializado que depende, acima de tudo, do julgamento subjetivo de profissionais treinados (CAOJETTE *et al.*, 1999). Portanto, de acordo com esses autores, ao analisar um crédito, faz-se uso de técnicas subjetivas e objetivas, mostrando que uma não substitui a outra.

Caouette *et al.* (1999) destacam que os modelos atuais de avaliação de risco de crédito estão mais para esforços pioneiros na busca de melhores opções do que

para a solução final. Alguns dos resultados destes esforços podem vir a ser completamente descartados, mas a maioria será incorporada a modelos que ainda serão construídos. Neste sentido, todos os modelos de análise de crédito são pontes para o futuro.

2.1.3.2 Crédito a Pessoas Físicas

De acordo com Silva (2006), o mercado de crédito à pessoa física de baixa renda no Brasil apresenta características paradoxais. De um lado, grande contingente de indivíduos recebe uma renda que compromete o equilíbrio orçamentário básico. De outro, muitas empresas oferecem promessa de crédito fácil e desburocratizado.

Segundo Schrickel (1997), o número mais expressivo de clientes que solicitam crédito a instituições financeiras são pessoas físicas. O crédito a pessoas físicas vem crescendo continuamente nos últimos anos e a taxas superiores quando comparado ao crédito a pessoas jurídicas. Entre dezembro de 2003 e julho de 2007, os empréstimos a pessoas físicas cresceram 179,9%, sendo que no mesmo período o aumento observado nas operações a pessoas jurídicas foi de 87,8%, como é apresentado na Figura 2.

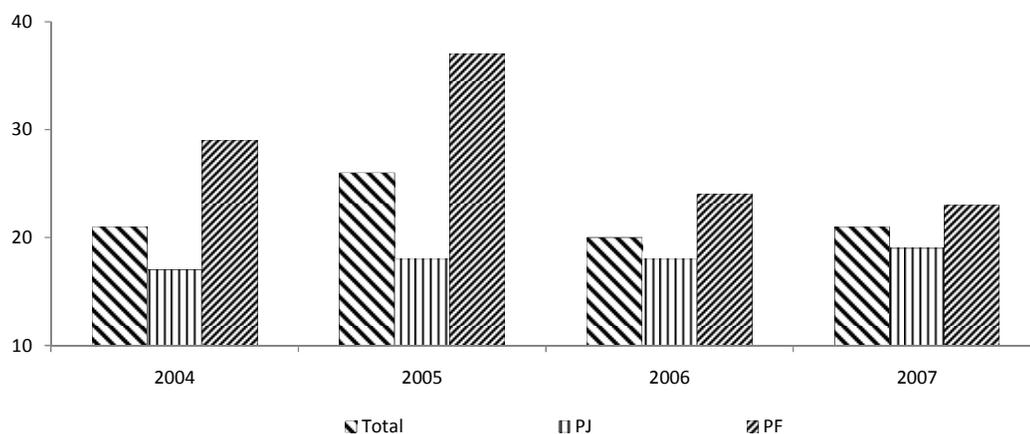


Figura 2 Taxas de crescimento do crédito (variação em doze meses)

Fonte: Banco Central do Brasil (2007, p. 41)

Santos (2000) comenta a respeito das linhas de crédito destinadas a pessoas físicas, que têm o objetivo de atender a três necessidades básicas:

- empréstimos emergenciais – atendem basicamente a necessidades imediatas do cliente, principalmente em situações de desequilíbrios orçamentários. São operações de curtíssimo prazo (inferior a um mês), sendo amortizado em uma única data de vencimento;
- financiamentos de compras – permitem atender às necessidades de aquisição de produtos e serviços para consumo e bem estar do cliente. Estes créditos são de curto prazo (inferior a doze meses), sendo que a amortização é geralmente efetuada mensalmente;
- investimentos – permitem ao cliente adquirir bens ou mesmo para o desenvolvimento de suas atividades profissionais. São operações de maior valor que geralmente são de longo prazo (superior a doze meses), com amortização parcelada.

De acordo com Schrickel (1997), transações de crédito a pessoas físicas envolvem uma grande quantidade de operações, onerando a estrutura de pessoal e os custos da instituição. Porém, esta característica também produz uma considerável diluição do risco, já que cada financiamento tem valor reduzido, quando comparados aos créditos concedidos a pessoas jurídicas. Outras características que diferenciam o crédito a pessoas físicas e jurídicas são apresentadas na Figura 3.

Pessoas FÍSICAS	Pessoas JURÍDICAS
<ul style="list-style-type: none"> ❖ grande quantidade de empréstimos ❖ baixos valores unitários ❖ baixo risco unitário ❖ alta margem unitária ❖ aprovação de crédito massificada, consoante certos conceitos gerais (renda, patrimônio), envolvendo a média administração ❖ tomadores fornecem dados financeiros limitados e particularizados ❖ retrospecto sobre o conceito de crédito restrito à emissão de cheques sem fundos ❖ administração de crédito na base de portfólio (estatísticas), revisões feitas na base caso a caso (por empréstimo) 	<ul style="list-style-type: none"> ❖ pequena quantidade de empréstimos ❖ altos valores unitários ❖ alto risco unitário ❖ baixa margem unitária ❖ aprovação de crédito seletiva (caso a caso pelo tomador, envolvendo os mais altos escalões da hierarquia) ❖ tomadores possuem demonstrações de maior credibilidade (auditadas) ❖ conceito sobre práticas comerciais e hábito de pagamentos rastreável e mais facilmente comprovável ❖ administração de crédito na base caso a caso (por tomador) com freqüente processo de revisão do limite de crédito em termos globais

Figura 3 Contraste entre crédito a pessoas físicas e pessoas jurídicas

Fonte: Schrickel (1997, p. 160)

De posse dessas informações, caberá a cada instituição definir o perfil de clientes que desejaria ter, bem como a metodologia de análise de crédito para cada caso (SCHRICHEL, 1997).

2.1.3.3 Os Cs do Crédito

Na análise de crédito subjetiva aplicada a solicitações de pessoas físicas, o objetivo é avaliar o risco da operação. Para isso, o analista toma como referência um conjunto de fatores denominados de Cs do Crédito, que são cinco palavras iniciadas com a letra C (SCHRICHEL, 1997). Para Silva, J. (2003), os Cs do crédito contêm as variáveis relacionadas ao risco do cliente e que poderão fornecer a base para a classificação do risco da operação.

De acordo com Schrickel (1997), decisões de concessão de crédito não devem ser tomadas levando em consideração apenas um dos Cs isoladamente. Na Figura 4 é apresentado um quadro com a caracterização conjunta dos 5 Cs do crédito apresentado por Santos (2000).

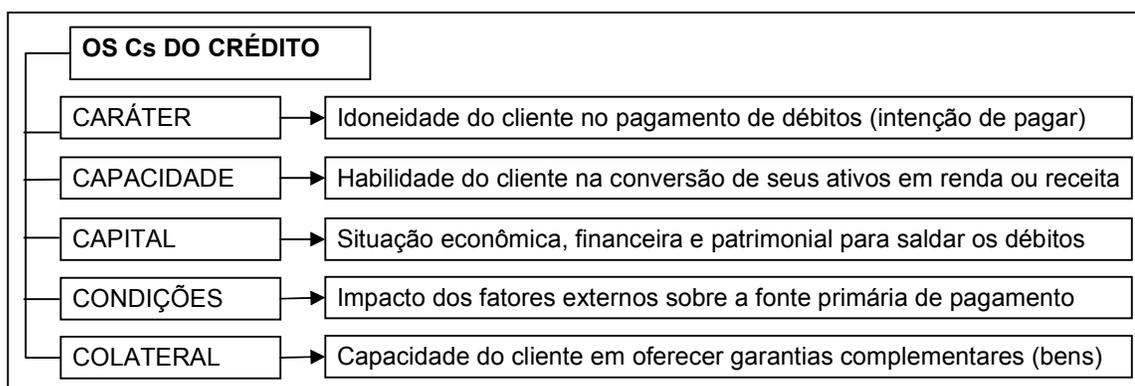


Figura 4 Quadro resumo dos Cs do crédito

Fonte: adaptado de Santos (2000, p. 45)

Para Caouette *et al.* (1999), a análise da capacidade financeira do cliente pode ser comparada à atividade de um alfaiate, sendo feita sob medida para as características do comprador. Ainda segundo esses autores, ao longo do tempo, as

técnicas analíticas de avaliação do risco de crédito evoluíram muito, permanecendo ainda os três Cs: caráter, capacidade e capital, como o tripé do crédito.

2.1.3.3.1 Caráter

Segundo Schrickel (1997), este é o critério mais importante e crítico, não importando sequer qual o valor da transação. Neste momento é feita uma minuciosa investigação a respeito dos antecedentes do cliente para análise da capacidade de pagamento. Esta busca por informações se dá através do preenchimento de uma ficha cadastral do tomador. Apesar de a ficha cadastral ser uma fonte valiosa de informações objetivas, ela não fornece todos os requisitos que permitem avaliar, subjetivamente, a real intenção do cliente em pagar seus débitos.

Silva, J. (2003) define caráter como a intenção de cumprir a promessa de pagamento. Cabe enfatizar que um indivíduo pode atrasar, ou até mesmo deixar de pagar, por não dispor de recursos, o que não é decorrência necessariamente de seu caráter. Isto mostra o quanto é difícil identificar o conjunto de boas e más qualidades intrínsecas à pessoa do proponente, face ao hábito de pagamento de suas contas. O autor ressalta ainda que o fator tempo pode demonstrar o quanto o caráter é mutável, de tal forma que, alguém é honesto até o dia em que deixa de ser. De qualquer forma, dados relativos ao passado do tomador podem ser instrumentos úteis para a tomada de decisão.

Para a análise do caráter é indispensável a existência de informações históricas do cliente, que demonstrem a intenção e pontualidade no pagamento de dívidas. A idoneidade financeira do cliente é uma das primeiras informações analisadas. Se existirem informações negativas e/ou pendências que sejam relevantes para a instituição financeira, de acordo com suas políticas de crédito, a decisão é a recusa imediata do crédito (SANTOS, 2000).

Para ter uma boa noção do caráter do tomador, é preciso analisar o cenário em que atua, como também a conjuntura econômica, sempre vislumbrando possíveis riscos para o pagamento da dívida. A esse respeito, Schrickel (1997, p. 49) destaca que “o caráter é um conceito que transcende ao indivíduo, alcançando todo

o extrato social e econômico do qual ele faz parte (grupos sociais, regiões, nações). Neste sentido, as decisões de crédito devem levar em conta a integridade ética do conjunto de pessoas no qual o tomador potencial está inserido”.

2.1.3.3.2 Capacidade

Para Silva, J. (2003), a capacidade refere-se à habilidade do tomador de crédito de gerir estratégica e operacionalmente seus negócios. Schrickel (1997) ressalta que se um indivíduo revela capacidade para administrar seu negócio, de tal forma a fazê-lo prosperar e ter sucesso, ele estará, desta forma, demonstrando sua capacidade em saldar compromissos.

Capacidade e caráter se confundem em certos aspectos. Esta confusão acontece porque pode o tomador ser de todo honesto, ou seja, tem a inquestionável vontade (caráter) de cumprir com suas obrigações. Contudo, se sua habilidade (capacidade) em fazê-lo não for satisfatória, ou seja, se ele for incapaz de pagar, a vontade demonstrada perde seu valor (SCHRICKEL, 1997).

Segundo Santos (2000), análise de capacidade refere-se à avaliação subjetiva do analista quanto à habilidade do tomador no gerenciamento de seus negócios e sua conversão em renda ou receita. Em geral, os credores atribuem à renda a denominação de fonte primária de pagamento, sendo o principal referencial para possibilitar a análise da compatibilidade do empréstimo com a capacidade financeira do tomador.

Schrickel (1997) indica outro ângulo a ser considerado na análise de capacidade: a análise do currículo profissional do tomador. Um indivíduo que demonstre ao longo de sua carreira estabilidade em empregos, crescimento em cargos de responsabilidade e atingimento de resultados relevantes, estaria apresentando evidências de sua capacidade.

2.1.3.3.3 Capital

Para Silva, J. (2003), a análise do capital compreende investigar a situação financeira, econômica e patrimonial do tomador, levando em consideração a composição dos recursos, onde são aplicados e como são financiados. Nesse sentido, Santos (2000) ressalta que os credores sempre enfrentam dificuldades para analisar a real ou aproximada situação financeira de seus clientes.

A análise financeira meticulosa é essencial para a decisão de crédito. Porém, ter acesso às informações que reflitam a exata situação financeira e patrimonial do tomador não é uma tarefa fácil. Tal fato se deve à possibilidade dos clientes manipularem ou omitirem informações com o objetivo de conseguirem aprovação de suas solicitações de crédito (SANTOS, 2000).

De acordo com Schrickel (1997), a figura do capital não fica tão evidente quando se trata de análise de crédito a pessoas físicas, já que os empréstimos pessoais não são respaldados em capital, mas na renda ou ganhos mensais do indivíduo. Para a análise de crédito destes clientes, prevalece quase sempre a figura do contracheque, afinal esta é a base primária para a análise do pagamento do empréstimo, e não a penhora de sua casa ou carro.

2.1.3.3.4 Condições

Para Schrickel (1997), as condições dizem respeito ao cenário micro e macroeconômico em que o tomador está inserido. Toda negociação de crédito deve levar em conta o contexto atual e as perspectivas futuras da economia. Desta forma, os credores tendem a ser mais cautelosos em momentos recessivos e mais liberais em momentos de recuperação econômica.

Segundo Santos (2000), a análise das condições está associada à avaliação do impacto de fatores externos sobre a fonte primária de pagamento. É de extrema importância para determinar o risco de crédito, já que dependendo da gravidade de situações como o aumento nas taxas de desemprego, o recebimento dos valores concedidos pode se tornar difícil.

A força do macroambiente pode manifestar-se tanto de forma positiva, representando oportunidades, quanto de modo negativo, trazendo ameaças. Dentre os fatores que podem influenciar estão planos econômicos, abertura da economia, política cambial e todo tipo de intervenções do governo na economia (SILVA, J. 2003).

2.1.3.3.5 Colateral

É grafada como em inglês e o seu significado é garantia, podendo ser encarada como uma espécie de 'ponto-socorro' para o credor. Na análise de uma solicitação de crédito, serve para contrabalançar e atenuar impactos negativos que possam ocorrer em decorrência do enfraquecimento de um dos três elementos: capacidade, capital e condições. Tal enfraquecimento implica em maior risco e o colateral serve para compensar esta elevação do risco. Raramente (nunca idealmente), o colateral deve ser utilizado para compensar os pontos fracos dentro do critério caráter, porque quando a honestidade está em falta, o crédito incluirá riscos que não devem ser assumidos pelo credor (SCHRICKEL, 1997).

Silva, J. (2003) refere-se à dimensão colateral como a capacidade do indivíduo em oferecer garantias complementares. A garantia é uma espécie de segurança adicional, sendo que em alguns casos, a concessão do crédito precisará dela para compensar as fraquezas decorrentes de outros fatores de risco.

Segundo Santos (2000), o colateral está associado com a análise da riqueza patrimonial (bens móveis e imóveis). Em casos de perda da fonte primária de pagamento do débito, o colateral pode ser considerado como a possibilidade futura de vinculação ao contrato de crédito.

Neste sentido, os C's do crédito podem ser utilizados como ferramenta adicional à análise feita com uso de modelos de previsão de risco de crédito. A diferenciação desses modelos em relação aos métodos subjetivos de análise de crédito se dá, principalmente, pelo fato da seleção dos fatores-chave e seus respectivos pesos ser realizada através de processos estatísticos (ARAÚJO, 2006).

2.2 MODELOS DE PREVISÃO DE RISCO DE CRÉDITO

Com o rápido desenvolvimento da informática, a partir dos anos 70, os sistemas de pontuação de crédito baseados na abordagem estatística surgiram no negócio de financiamento a pessoas físicas e jurídicas. Tais sistemas são tidos como um dos métodos mais importantes de suporte à tomada de decisão para grandes volumes de solicitações de crédito (SANTOS, 2000).

Os modelos de previsão de risco de crédito vêm como uma ferramenta de auxílio ao crédito massificado, que é caracterizado pela avaliação de um grande número de solicitações de pequenos valores, já que a competitividade do mercado exige decisões rápidas. O analista informa os dados de seu potencial cliente no sistema de crédito e, imediatamente, o computador fornece a informação quanto à aprovação do crédito. O método estatístico utilizado para a construção do modelo leva em consideração o histórico da instituição com seus clientes, possibilitando a identificação das características capazes de diferenciar o bom do mau pagador (SILVA, 2006).

Alguns autores como Caouette *et al.* (1999) e Silva, J. (2003) têm citado a análise estatística multivariada como uma ferramenta poderosa na avaliação do risco de inadimplência presente na concessão de crédito. Segundo Silva (1988), uma das vantagens do uso de técnicas estatísticas para elaboração de sistemas de pontuação é que os pesos a serem atribuídos aos índices são determinados por cálculos e processos estatísticos, o que exclui a subjetividade no momento da análise ou mesmo o estado de espírito do analista do crédito.

Após a atribuição de valores numéricos a cada característica ou variável de risco selecionada do tomador, obtém-se uma pontuação, que determinará se o crédito pode ser concedido ou recusado de maneira padronizada, consistente e objetiva, baseando-se nas probabilidades de reembolso calculadas (SANTOS, 2000). Tal ponderação é construída, buscando uma separação dos clientes segundo seu desempenho de crédito.

Pereira *et al.* (2002) definem como créditos bons os clientes que nunca atrasaram ou no pior dos casos foram moderadamente inadimplentes. Os maus

créditos são aqueles em condições de atraso superior a 90 dias uma vez em um ano ou mais de 60 dias por duas vezes em um ano. Segundo o autor, definições mais complexas sobre mau crédito (quatro atrasos de 30 dias, três atrasos de 60 dias e um de 90 dias) são difíceis de acompanhar e não garantem um aumento substancial do poder de previsão do modelo.

A pontuação obtida com a utilização dos modelos de previsão pode ser usada de duas formas: interpretada literalmente como a probabilidade de ocorrência do não pagamento ou servir como uma medida de classificação para designar um potencial tomador em um grupo de bom ou mau cliente, comparando a pontuação obtida com um ponto de corte (SAUNDERS, 2000).

Nos últimos tempos, o uso de métodos para previsão de risco de crédito tem sido muito divulgado. Isto tem feito com que as instituições financeiras saiam numa corrida em busca dessas ferramentas. Contudo, esses métodos não podem ser entendidos como receitas milagrosas capazes de resolver todos os problemas relacionados ao risco de operações de crédito (SILVA, J. 2003). Segundo Caouette *et al.* (1999), a adoção de modelos que visam estabelecer o risco de crédito em relação ao tomador não tornará os negócios isentos de risco.

Caouette *et al.* (1999) afirmam que modelos de previsão que medem o risco de crédito ainda estão em etapa de desenvolvimento; desta forma são úteis, porém ainda imperfeitos. Os autores ainda fazem um alerta: se os modelos forem manipulados sem o cuidado devido e sem a consciência sobre seu uso como ferramentas de apoio à análise, podem aumentar, e não minimizar, a exposição de uma instituição ao risco de crédito.

Segundo Santos (2000), embora o uso de modelos de previsão seja de grande importância para a análise de crédito, há limitações para o seu uso: i) a determinação do risco de crédito deve considerar o impacto de fatores sistemáticos ou externos; ii) a utilização de informações imperfeitas dos tomadores e por dados disponíveis no mercado; iii) a ausência de banco de dados com informações dos tomadores em todo o mercado de crédito; e iv) as dificuldades de ajustamento de estratégias de diversificação em carteiras de empréstimos.

Mallister e Mingo (1994) também concordam que o grande problema para implantação de sistemas de pontuação está na indisponibilidade e defasagem de informações completas e verídicas dos tomadores. Muitas instituições financeiras têm interesse e, mais ainda, necessitam dos benefícios desses modelos, mas não têm um banco de dados com informações históricas de seus clientes, de forma a poder ponderar os fatores que levariam à obtenção da probabilidade de perdas com novos tomadores de crédito.

Silva, J. (2003) aponta outro limitador dos modelos que diz respeito ao tempo. Com o decorrer do tempo, tanto as variáveis quanto seus pesos relativos sofrem freqüentes alterações. Tal característica demonstra a necessidade de avaliações periódicas para avaliar o desempenho dos resultados de predição.

Na mensuração do risco de crédito, segundo Caouette *et al.* (1999), os modelos de pontuação podem ser classificados de acordo com três dimensões diferentes: as técnicas empregadas, o domínio de aplicações no processo de crédito e os produtos a que se aplicam.

As técnicas mais comumente empregadas são:

- técnicas econométricas: métodos estatísticos tais como regressão logística, análise discriminante, análises *logit* e *probit* modelam a probabilidade de inadimplência como variável dependente, explicada através de um conjunto de variáveis independentes;
- redes neurais: sistemas computacionais que tentam simular o funcionamento do cérebro humano por meio de uma rede de neurônios interligados. Uma alternativa às tradicionais técnicas estatísticas, por buscar a solução através de um processo de aprendizagem;
- modelos de otimização: técnicas de programação matemática que identificam os pesos ideais das variáveis, de forma a minimizar o erro do credor e maximizar seus lucros;
- sistemas especialistas: sistemas que imitam o processo usado por um analista experiente, que baseado em um conjunto de regras decisórias,

tentam reproduzir seu raciocínio estruturado, auxiliando, portanto, na tomada de decisão quanto à concessão do crédito;

- sistemas híbridos: sistemas movidos em parte por uma relação causal, cujos parâmetros são determinados por meio de técnicas de estimativa, utilizando computação, estimativas e simulação diretas.

Caouette *et al.* (1999) defendem que os sistemas de pontuação de risco de crédito são importantes por dispor ao credor o conhecimento que não estaria, de outra maneira, prontamente disponível. Em sua contribuição, ou autores ainda acrescentam que há uma grande vantagem competitiva com a utilização dos modelos, pois com um sistema de pontuação integrado é possível operar em diversas regiões geográficas, envolvendo diversas pessoas e, mesmo assim, operar com elevado grau de objetividade nas decisões.

Souza e Chaia (2000) ressaltam que apesar de os sistemas de pontuação representarem um processo científico, não eliminam a possibilidade de se recusar um bom cliente ou que se aceite um mau pagador. Isto se deve ao fato de que nenhum modelo de previsão de risco de crédito alcança o total de características relevantes para discriminar e classificar os tipos de clientes. Mesmo que isso fosse possível, o custo de obtenção do modelo tornaria a análise economicamente inviável.

Segundo Silva, J. (2003), o uso de modelos de pontuação presta uma grande contribuição ao crédito. Contudo, o uso desses métodos não elimina a necessidade de que as instituições financeiras tenham definições políticas e estratégicas claras e que seus profissionais estejam treinados para uma boa análise de crédito. Deste modo, os modelos devem ser entendidos como um instrumental complementar para o analista.

São diversas as técnicas a serem utilizadas para a construção dos modelos de previsão de risco de crédito. Dentre os métodos possíveis estão: regressão linear múltipla, programação linear, algoritmos genéticos, árvore de decisão, análise discriminante, regressão logística, redes neurais e, mais recentemente, a análise de

sobrevivência como exposto no trabalho de Andreeva *et al.* (2007). Na seqüência são descritas as três técnicas multivariadas utilizadas neste trabalho.

2.2.1 Análise Discriminante

A análise discriminante envolve a determinação de uma combinação linear de variáveis independentes que discriminarão melhor uma observação entre grupos definidos *a priori* (HAIR *et al.*, 2005). Segundo Johnson e Wichern (2002), a análise discriminante é uma técnica multivariada com o objetivo de tratar dos problemas que envolvem separar conjuntos distintos e alocar novos objetos em conjuntos previamente definidos.

Silva, J. (2003) acrescenta que a análise discriminante é uma ferramenta estatística utilizada para classificar um elemento em um determinado grupo. Para isso, é necessário que o elemento a ser classificado pertença realmente a um dos grupos em estudo, e que sejam conhecidas as características dos elementos dos grupos, permitindo a comparação entre as características do elemento que se deseja classificar com as características dos elementos dos diversos grupos.

Conforme Johnson e Wichern (2002), a idéia de discriminar e classificar foi introduzida por Ronald. A. Fisher no primeiro tratamento moderno dos problemas de separação de conjuntos em seu trabalho sobre espécies de plantas em 1935. O método proposto por Fisher consiste basicamente em separar um conjunto de objetos em duas classes pré-definidas. A idéia de Fisher foi transformar as observações multivariadas X em observações univariadas Y 's, de tal modo que fossem separadas tanto quanto possível. Para obter essa separação, é proposto o uso de combinações lineares das variáveis independentes que compõem X para obtenção do valor previsto de Y .

Vários estudos deram seqüência para a utilização desta técnica em diversas áreas. Em Finanças, um dos mais relevantes estudos na previsão do risco de crédito foi realizado por Edwards I. Altman, em 1968. A partir do artigo "*Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*", o autor seguiu com estudos sobre a capacidade preditiva de índices oriundos de demonstrações financeiras das empresas (CORRAR *et al.*, 2007).

Caouette *et al.* (1999) destacam que embora não seja tão difundida quanto outras técnicas estatísticas, como por exemplo a regressão linear, a análise discriminante tem tido crescente utilização nas áreas de economia e finanças. Muitas das aplicações consistem na construção de modelos destinados à previsão de falência de empresas e à inadimplência de pessoas físicas.

Corrar *et al.* (2007) ressaltam que as variáveis independentes geralmente são métricas com valores contínuos, mas também podem assumir valores que representam categorias (alto e baixo, por exemplo). Na área de Marketing há estudos que se encaixam nesta última situação, porém os autores destacam que técnicas como regressão logística e redes neurais são mais apropriadas a esse tipo de variável independente em razão das características de seus algoritmos.

De acordo com Hair *et al.* (2005), a discriminação é alcançada estabelecendo-se os pesos discriminantes, que são calculados com o objetivo de maximizar a variância entre os grupos e, por consequência, minimizar a variância dentro dos grupos. A combinação linear para uma análise discriminante, também conhecida como função discriminante, tem a forma apresentada na Equação 01.

$$Z_k = a + W_1X_{1k} + W_2X_{2k} + \dots + W_nX_{nk} \quad (01)$$

onde,

Z_k = escore Z discriminante para o objeto k

a = intercepto

W_i = coeficiente discriminante para a variável independente i

X_{ik} = variável independente i para o objeto k

O processo de decisão para aplicação da análise discriminante pode ser vista como a construção de um modelo que envolve 7 estágios, conforme é apresentado na Figura 5. O primeiro passo na análise é o estabelecimento dos objetivos, seguido da definição das questões específicas de planejamento e da certificação do atendimento das suposições inerentes ao uso da técnica. A seqüência da análise se dá com a dedução da função discriminante e verificação de sua significância estatística. Os resultados da aplicação da função são avaliados para verificar a precisão preditiva, juntamente com a interpretação da função discriminante, identificando quais as variáveis independentes que mais contribuem

para a separação dos grupos. Por fim a função discriminante é validada com uma amostra de teste.

Estágio 1 – Problema de Pesquisa e Objetivos	<ul style="list-style-type: none"> • Avaliar diferenças de grupos em um perfil multivariado • Classificar observações em grupos • Identificar dimensões de discriminação entre grupos
Estágio 2 – Questões de Planejamento de Pesquisa	<ul style="list-style-type: none"> • Seleção das variáveis independentes • Considerações sobre o tamanho da amostra • Criação de amostras de análise e de teste
Estágio 3 – Suposições	<ul style="list-style-type: none"> • Normalidade de Variáveis Independentes • Linearidade de relações • Falta de multicolinearidade entre variáveis independentes • Matrizes de dispersão iguais
Estágio 4 – Estimação da Função Discriminante	<ul style="list-style-type: none"> • Estimação simultânea ou <i>stepwise</i> • Significância da função discriminante
Estágio 5 – Avaliação da Precisão Preditiva com Matrizes de Classificação	<ul style="list-style-type: none"> • Determinar escore de corte ótimo • Especificar critério para avaliar razão de sucesso • Significância estatística de precisão preditiva
Estágio 6 – Interpretação da Função Discriminante	<ul style="list-style-type: none"> • Pesos discriminantes • Cargas discriminantes • Valores <i>F</i> parciais
Estágio 7 – Validação dos Resultados Discriminantes	<ul style="list-style-type: none"> • Subamostras ou validação cruzada • Perfil de diferenças de grupos

Figura 5 Estágios de decisão da análise discriminante

Fonte: adaptado de Hair *et al.* (2005, p. 217 e 218)

Segundo Hair *et al.* (2005), no estágio de definição dos objetivos, cabe ao pesquisador delimitar seu problema de pesquisa entre as opções: (i) determinar diferenças entre os grupos, identificando quais variáveis independentes têm maior poder de explicação; (ii) classificar objetos em grupos com base em seus escores em um conjunto de variáveis independentes e (iii) estabelecer o número e a composição das dimensões de discriminação entre grupos formados a partir das variáveis independentes.

A questão do tamanho de amostra é bastante relevante, visto que a análise discriminante é muito sensível à proporção entre tamanho da amostra e o número de variáveis independentes. Muitos estudos sugerem a utilização de uma proporção de 20 observações para cada variável em estudo. Outro ponto quanto à amostra refere-se à divisão em amostra de análise ou estimação e amostra de teste ou validação. É essencial que cada subamostra tenha tamanho adequado para suportar as conclusões dos resultados. No entanto nenhuma regra rígida foi estabelecida, e

alguns autores preferem partições na proporção 60-40 ou 75-25 entre os grupos de análise e teste, respectivamente (HAIR *et al.*, 2005).

A normalidade das variáveis independentes e a igualdade das matrizes de variância e covariância dos grupos são as suposições básicas a serem atendidas para possibilitar a utilização da análise discriminante. Outra característica dos dados que pode afetar os resultados é a multicolinearidade entre as variáveis independentes (HAIR *et al.*, 2005). Violações graves dessas suposições poderão prejudicar a eficiência do modelo e, portanto, a extrapolação dos seus resultados para a população (JOHNSON; WICHERN, 2002).

Segundo Guimarães e Chaves Neto (2002), uma das principais ressalvas ao uso da análise discriminante diz respeito à natureza das variáveis estudadas. Todas as suposições necessárias ao desenvolvimento da técnica tratam as variáveis como contínuas, quando na prática é comum o uso de variáveis categóricas ou ordinais.

Segundo Hair *et al.* (2005), dois métodos computacionais podem ser utilizados para determinar uma função discriminante: o método simultâneo (direto) e o método *stepwise*. A estimação *stepwise* é uma alternativa à abordagem simultânea, pois envolve a inclusão das variáveis independentes na função discriminante, uma por vez, com base em seu poder discriminatório. Com a função discriminante determinada, deve-se testar sua significância. Entre os testes estatísticos mais utilizados para avaliar o poder discriminatório estão: lambda de Wilks, traço de Hotelling e o critério de Pillai.

Segundo Corrar *et al.* (2007), o método *stepwise* é considerado como uma das ações corretivas para os problemas de multicolinearidade. O procedimento de avaliação das variáveis independentes desconsidera variáveis que apresentem sinais de multicolinearidade, optando por manter no modelo a de maior significância estatística.

Para testar a significância estatística da função discriminante, Hair *et al.* (2005) sugerem ainda a utilização de uma medida generalizada da distância entre os centróides dos grupos. Essa medida compara as distribuições dos escores discriminantes dos dois grupos. Quando a sobreposição das distribuições é

pequena, considera-se que a função separa bem os grupos, mas se a sobreposição é grande, a função discriminante tem uma separação pobre dos grupos. Na Figura 6 é melhor ilustrado esse conceito, utilizando duas distribuições de escores.

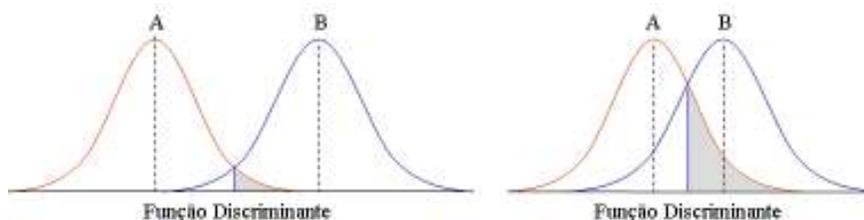


Figura 6 Representação de escores Z discriminantes

Fonte: Hair *et al.* (2005, p. 209)

A avaliação da precisão preditiva com matrizes de classificação, proposta no estágio 5, é comentada em detalhes na seção 2.2.4.

No estágio 6, é feito o exame da função discriminante para determinar a importância relativa de cada variável independente na discriminação entre os grupos. Hair *et al.* (2005) sugerem a utilização de três métodos para determinar essa importância: (i) pesos discriminantes padronizados, (ii) cargas discriminantes (correlações de estrutura) e (iii) valores F parciais.

Os pesos discriminantes representam a contribuição relativa da variável independente e o sinal indica se a variável tem uma contribuição positiva ou negativa. As cargas discriminantes medem a correlação linear simples entre cada variável independente e a função discriminante (HAIR *et al.*, 2005). Para verificar se a discriminação é boa ou não, Fischer sugere fazer uma análise de variância, através do teste F , para verificar a significância do poder discriminador das variáveis independentes consideradas (SILVA, J., 2003).

O sétimo e último estágio, de validação do modelo estimado, é essencial para garantir que os resultados tenham validade interna e externa, o que é feito através da amostra de teste previamente separada para este fim. Segundo Hair *et al.* (2005), esse procedimento se justifica por eliminar um viés ascendente que ocorrerá na predição se os indivíduos classificados na etapa de validação forem os mesmos

da amostra de análise. Desta forma, a validação do modelo estaria simulando a utilização da função discriminante na classificação de um novo indivíduo.

2.2.2 Regressão Logística

A utilização da técnica de regressão logística é adequada em muitas situações porque permite que se analise o efeito de uma ou mais variáveis independentes (discretas ou contínuas) sobre uma variável dependente dicotômica, representando a presença (1) ou ausência (0) de uma característica (HOSMER; LEMESHOW, 1989).

A técnica de regressão logística foi desenvolvida por volta de 1960 em resposta ao desafio de realizar previsões ou explicar a ocorrência de determinados fenômenos em que a variável dependente fosse de natureza binária. Um dos estudos pioneiros que mais contribuíram para o avanço da técnica foi o famoso *Framingham Heart Study*, realizado com a colaboração da Universidade de Boston. O objetivo principal do estudo foi identificar os fatores que contribuíam para a ocorrência de doenças cardiovasculares (CORRAR *et al.*, 2007).

De acordo com Hosmer e Lemeshow (1989), a regressão logística tornou-se, portanto, um método padrão de análise de regressão para variáveis medidas de forma dicotômica. Desta forma, a diferença principal da regressão logística quando comparada ao modelo linear clássico é que a distribuição da variável resposta segue uma distribuição binomial, e não uma distribuição normal.

A esse respeito, Hair *et al.* (2005) afirmam que a regressão logística se assemelha em muitos pontos à regressão linear, mas se difere basicamente no sentido de prever a probabilidade de um evento ocorrer. Para obter um valor previsto delimitado entre zero e um, usa-se uma relação assumida entre as variáveis independentes e a variável dependente que lembra uma curva em forma de 'S', a distribuição sigmóide, representada na Figura 7.

Os modelos lineares de regressão não podem acomodar tal relação entre as variáveis, já que ela é inerentemente não-linear. Por isso a regressão logística foi desenvolvida para lidar especificamente com essas questões. A regressão logística

deriva seu nome justamente dessa transformação logística utilizada com a variável dependente (HAIR *et al.*, 2005).

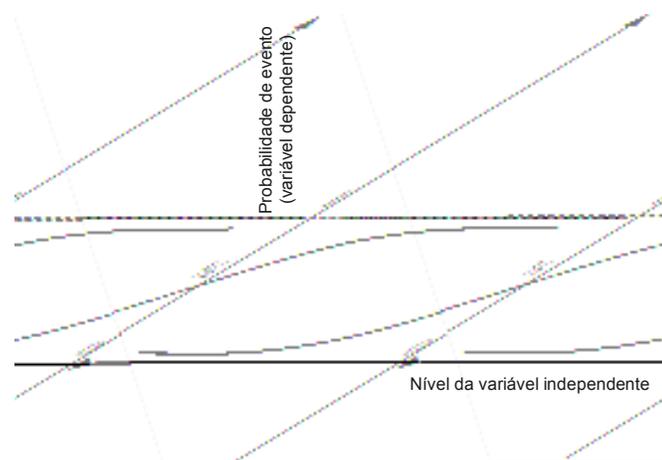


Figura 7 Forma da relação logística entre as variáveis

Fonte: Hair *et al.* (2005, p. 232)

Como o uso do modelo linear poderia conduzir a predições de valores menores que zero e maiores que um, torna-se necessário converter as observações em razão de chances (*odds ratio*) e submetê-las a uma transformação logarítmica. Com isso, o modelo passa a evidenciar mudanças nas inter-relações dos *logs* da variável independente (CORRAR *et al.*, 2007).

Segundo Hair *et al.* (2005), devido à natureza não-linear da transformação logística, para a estimação do modelo é necessária a utilização de um procedimento que, de forma iterativa, encontra as melhores estimativas para os coeficientes: o método da máxima verossimilhança. Isso resulta na análise do valor de verossimilhança ao invés da soma de quadrados, utilizada na regressão linear, como medida do ajuste geral do modelo.

O modelo de regressão logística é obtido pelo procedimento de comparação da probabilidade de um evento ocorrer com a probabilidade de não ocorrer. De acordo com Hair *et al.* (2005), esta razão pode ser expressa segundo a Equação 02.

$$\frac{\text{Prob(evento ocorrer)}}{\text{Prob(evento não ocorrer)}} = e^{B_0 + B_1 X_1 + \dots + B_n X_n} \quad (02)$$

Os coeficientes estimados (B_0, B_1, \dots, B_n) são medidas das variações na proporção das probabilidades, chamada de razão de desigualdade. São expressos em logaritmos, necessitando serem transformados para facilitar a interpretação. Um coeficiente positivo revela que aquela variável aumenta a probabilidade de ocorrência do evento, enquanto que um valor negativo diminui a probabilidade prevista.

Ao utilizar a técnica de regressão logística, o interesse pode estar na identificação do efeito de um fator de risco específico ou em determinar quais são os vários fatores associados com a variável resposta. Segundo Hosmer e Lemeshow (1989), a função logística vem sendo bastante utilizada não apenas pela simplicidade de suas propriedades teóricas, mas, principalmente, devido a sua simples interpretação como o logaritmo da razão de chances (*odds ratio*).

Para testar a significância dos coeficientes, Hair *et al.* (2005) sugerem o uso da estatística de Wald. Ela fornece a significância estatística para cada coeficiente estimado, de modo que o teste de hipóteses pode ocorrer como acontece na regressão múltipla.

Para Hair *et al.* (2005), outra semelhança com a regressão múltipla é o fato de que dados nominais e categóricos podem ser tomados como variáveis independentes do modelo por meio de codificação dicotômica. Além disso, os procedimentos de seleção de modelos *stepwise forward* e *stepwise backward* também estão disponíveis.

Segundo Corrar *et al.* (2007), um dos motivos pelos quais a regressão logística tem sido muito utilizada é o pequeno número de suposições. Com esta técnica, o pesquisador consegue contornar certas restrições encontradas em outros modelos multivariados.

A regressão logística não depende de suposições rígidas, tais como a normalidade das variáveis independentes e a igualdade das matrizes de covariância nos grupos. Essas suposições geralmente não são válidas em muitas situações práticas, principalmente quando há variáveis independentes de natureza não

métrica. Portanto, quando esses pressupostos não são satisfeitos, a aplicação da regressão logística ainda se torna apropriada (HAIR *et al.*, 2005).

Mesmo com toda sua flexibilidade, existe o pressuposto importante da alta correlação entre as variáveis independentes, já que o modelo de regressão logística é sensível à colinearidade entre as variáveis (HAIR *et al.*, 2005). A utilização de variáveis altamente correlacionadas para a estimação do modelo pode ocasionar estimativas extremamente inflacionadas dos coeficientes de regressão (HOSMER; LEMESHOW, 1989).

Portanto, o pesquisador que tem um problema que envolva uma variável dependente dicotômica não precisa apelar para métodos elaborados para suprir as limitações da regressão múltipla, nem precisa forçar-se a usar a análise discriminante, principalmente se suas suposições estatísticas não são satisfeitas. A regressão logística aborda satisfatoriamente esses problemas e oferece um método de análise desenvolvido especialmente para lidar com esse tipo de situação da forma mais eficiente possível (HAIR *et al.*, 2005).

2.2.3 Redes Neurais

Redes neurais é uma das técnicas de tratamento de dados mais recentes e que tem despertado grande interesse tanto de pesquisadores da área de tecnologia quanto da área de negócios (CORRAR *et al.*, 2007).

Pareciam ser muito promissoras quando foram iniciadas em meados do século XX, porém no final dos anos 60, pesquisas revelaram que as redes neurais daquela época eram de fato muito limitadas em capacidade, fazendo com que a área sofresse um retrocesso. Na década de 80, com o avanço da informática, o interesse recomeçou, dando continuidade aos estudos da área (HAIR *et al.*, 2005).

As redes neurais vêm sendo aplicadas mais recentemente na obtenção de modelos de pontuação para risco de crédito (SANTOS, 2000). Segundo Corrar *et al.* (2007), nesta área de negócios, as redes neurais têm encontrado várias aplicações interessantes com resultados superiores aos métodos estatísticos convencionais.

Estão complementando e enriquecendo técnicas estatísticas e sistemas especialistas.

De acordo com Hair *et al.* (2005), é uma abordagem totalmente diferente em relação a qualquer outra técnica multivariada. A diferença não está somente na estrutura, mas também no processo, já que as redes neurais têm um elemento-chave: a aprendizagem. Essa é outra analogia com o cérebro humano, através da qual, erros de saída são retornados ao início da rede, sendo ajustados adequadamente.

Haykin (2001, p. 28) acrescenta que “uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso”.

A estrutura e operação das redes neurais podem ser descritas por quatro conceitos: (i) o tipo de modelo de rede neural (ii) as unidades de processamentos (nós) que coletam informações, processam e criam um valor de saída; (iii) o sistema de nós arranjados para transferir sinais dos nós de entrada para os nós de saída, através dos nós intermediários; e (iv) a função de aprendizado pela qual o sistema ‘retorna’ erros na previsão para ajustar o modelo (HAIR *et al.*, 2005).

Haykin (2001) apresenta o elemento mais básico de uma rede neural, representado. O nó é análogo ao neurônio do cérebro humano, recebendo informações de entrada e criando resultados de saída. O processamento dessa informação acontece pela criação de um valor somado no qual cada entrada é multiplicada por seu respectivo peso. Esse valor é então processado por uma função de ativação, gerando uma saída que é enviada para o nó seguinte. Em geral, a função de ativação é não-linear, como a função sigmóide, da classe geral de curvas em forma de ‘S’ que incluiu a função logística. Outro elemento de entrada dos nós, chamado bias, utilizado também no processamento, funciona como uma constante da função aditiva. A representação de um neurônio artificial é apresentada na Figura 8 e seus elementos são descritos no Anexo A.

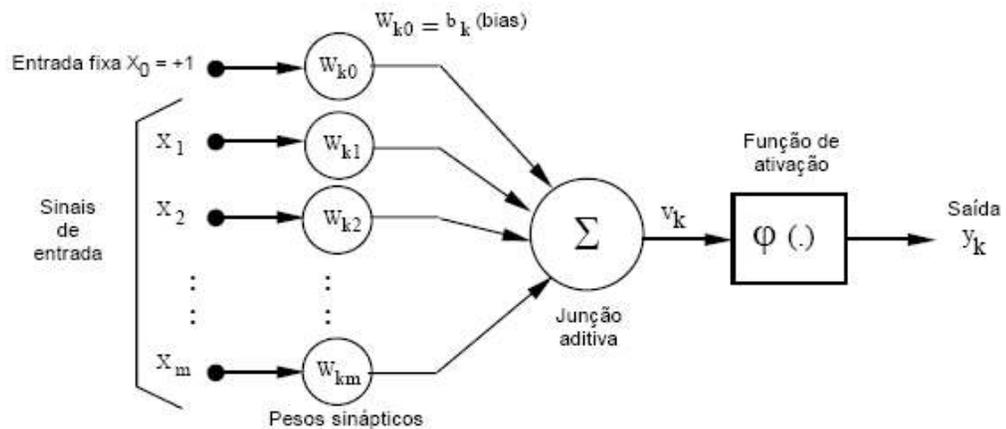


Figura 8 Modelo não-linear de um neurônio

Fonte: Haykin (2001, p. 36)

Uma rede neural é um arranjo seqüencial de três tipos de nós: de entrada, de saída e intermediários (ocultos ou escondidos). Os nós de entrada recebem os dados de cada caso e os transmitem para o restante da rede. Variáveis métricas necessitam apenas um nó para cada variável, já variáveis não-métricas precisam ser codificadas, de forma que cada categoria é representada por uma variável binária (HAIR *et al.*, 2005). Na Figura 9 é apresentado um modelo representativo do arranjo de uma rede neural.

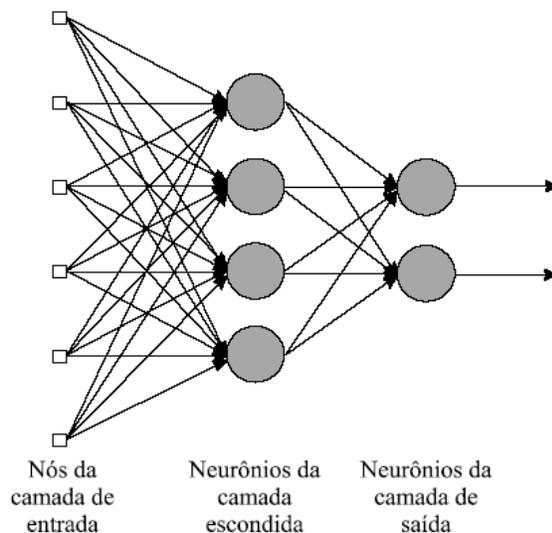


Figura 9 Modelo de uma rede neural

Fonte: Haykin (2001, p. 49)

Segundo Hair *et al.* (2005), um nó de saída recebe entradas e obtém um valor de saída, sendo este o resultado da previsão. É através das camadas ocultas e da função de ativação que a rede neural consegue representar as relações não-lineares entre as variáveis.

O tipo de problema a ser analisado pela rede neural pode ser restringido por sua arquitetura. Redes multicamadas são apropriadas para análise de problemas mais complexos, podendo ser considerada qualquer função contínua, linear ou não linear. Já as redes sem camadas escondidas conseguem fazer apenas separações lineares (CORRAR *et al.*, 2007). Neste sentido, Oliveira (2004) fez vários testes de arquitetura de redes neurais. Os resultados indicaram não haver melhoras consideráveis nos modelos com mais de uma camada escondida.

As redes neurais não têm a capacidade de transformar dados de qualidade ruim ou mal coletados em um modelo bem sucedido. Portanto, é essencial examinar os dados cuidadosamente, assim como em qualquer outro método estatístico. Para usar um modelo de rede neural, é necessária a preparação dos dados, devendo-se considerar o tamanho da amostra e distribuições dos dados (HAIR *et al.*, 2005).

Em termos particularmente simples, Haykin (2001) apresenta as fases necessárias à implantação de uma rede neural:

- a. organizar uma base de dados com toda a informação relevante e disponível;
- b. separar esta informação em dois grupos, sendo o primeiro para a fase de aprendizagem da rede e o segundo para a fase de teste;
- c. fazer as transformações e codificações necessárias nas variáveis de modo a servir de entrada para a rede;
- d. selecionar, treinar e testar a rede – esta fase deverá ser repetida até que os resultados sejam considerados adequados;
- e. aplicar o modelo desenvolvido ao problema de pesquisa.

Segundo Corrar *et al.* (2007), um dos parâmetros que mais impactam o tempo de aprendizado é o número de variáveis independentes, e por isso é muito importante que elas sejam selecionadas de forma criteriosa, a fim de que se utilizem aquelas que possam contribuir para a capacidade preditiva da rede. A separação do conjunto de dados em amostra de análise e teste é normalmente feita de forma automática pelo *software*. Dependendo do *software* utilizado, é necessária uma padronização nos dados, de forma que os valores estejam entre zero e um, garantindo um melhor processamento da rede.

Após a preparação dos dados, é necessário então pensar na estrutura da rede neural, determinando o número de camadas e neurônios e o método de aprendizado. Na área de negócios, costuma-se utilizar uma única camada intermediária, mas o número de neurônios nessa camada pode variar bastante. Na prática, o que se faz é testar até encontrar a melhor rede por tentativa e erro. Esta etapa é muito laboriosa, dado o grande número de combinações de parâmetros a serem testados. Atualmente, alguns *softwares* mais sofisticados possuem a capacidade de facilitar essa tarefa, fazendo de forma automática a geração e teste das redes (CORRAR *et al.*, 2007).

Hair *et al.* (2005) afirmam que a característica que realmente destaca as redes neurais em relação às outras técnicas estatísticas é sua capacidade de 'aprender' ou de 'corrigir a si mesma'. O objetivo é analisar um grande número de casos pela rede neural na fase de treinamento de forma que ela possa fazer as melhores previsões.

A forma mais comum de treino da rede é a retropropagação que consiste basicamente de dois passos: um passo para frente, a propagação, e um passo para trás, a retropropagação. Na propagação, as variáveis de entrada são apresentadas aos nós e seu efeito se propaga através da rede, camada por camada, obtendo-se uma saída para a rede, sem alteração dos pesos sinápticos. O valor do erro é calculado, comparando-se a saída da rede com a saída esperada. Na retropropagação, os pesos sinápticos são ajustados de acordo com valor do erro, tentando reduzir esse erro e ajustando a rede para que a resposta seja mais próxima possível da resposta esperada (HAYKIN, 2001).

Depois de ter passado por um número suficiente de exemplos da amostra de análise, os pesos na rede não tem mais alterações significativas e o erro não se reduz mais. É neste ponto que o treinamento da rede é encerrado, ou seja, a rede aprendeu a reconhecer as entradas (CORRAR *et al.*, 2007).

Saunders (2000) salienta que na intenção de melhorar a predição do modelo, muitos pesquisadores usam uma grande rede neural com várias camadas para incluir grandes transformações nas combinações das variáveis independentes. Porém, com isso aparece o problema do ‘excesso de encaixe’, ou seja, obtenção de um modelo que explica quase com erro zero para a amostra de desenvolvimento, mas pode apresentar desempenho muito baixo na previsão com dados de uma amostra de validação.

Hair *et al.* (2005) apresentam duas maneiras de evitar o super ajuste da rede. A primeira é através de um limite mínimo de erro estabelecido que determinará quando o treinamento da rede deve parar. Os autores alertam que, apesar de arbitrário, essa abordagem diretamente evita o problema. Um segundo método é monitorar a taxa de erro para as amostras de estimação e de validação. Em algum ponto, o erro da amostra de validação vai se estabilizar e depois começar a aumentar. Esse é o ponto no qual a amostra de estimação começa a ficar super-treinada, se tornando muito particularizada à amostra, não sendo generalizável para a população. A representação da busca por este ponto de parada no treinamento da rede é ilustrada na Figura 10.

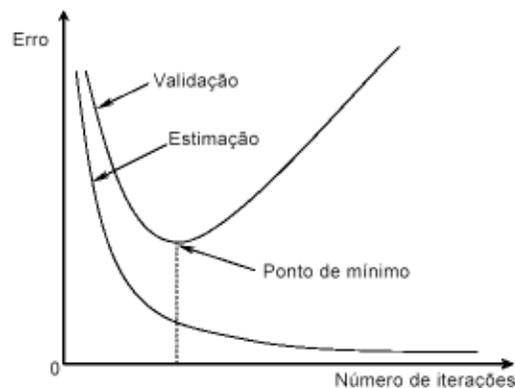


Figura 10 Treinamento versus capacidade de generalização da rede neural

Fonte: Haykin (2001, p. 243)

Smith (1996) sugere ainda que o super-ajuste da rede pode também ser evitado com o aumento da amostra de estimação, redução do número de neurônios na camada intermediária e do número de iterações para treinamento da rede.

Depois de treinada, uma rede neural está apta a associar um conjunto de valores que são apresentados em suas entradas a um resultado de saída. Porém, esta capacidade não é apenas uma memória, ela tem a habilidade da generalização. Uma rede neural pode descobrir respostas corretas mesmo nos casos em que os dados estão incompletos ou danificados (HAYKIN, 2001).

Ainda que o modelo de rede neural possa ser utilizado em situações que outras técnicas estatísticas, tais como regressão múltipla, análise discriminante e regressão logística, seriam também indicadas, ele não informa sobre a importância relativa das variáveis independentes na predição devido à combinação não-linear de pesos que ocorre na camada oculta. Nesse contexto, Hair *et al.* (2005) indicam a aplicação de redes neurais em problemas de previsão e classificação quando o interesse está na precisão de classificação e não na interpretação da variável estatística. Outro ponto negativo apontado por Corrar *et al.* (2007) diz respeito à escolha da arquitetura da rede que é por tentativa e erro. Tal característica pode tornar o processo muito demorado.

Saunders (2000) argumenta que a aplicação de técnicas não-lineares, como redes neurais, à análise de risco de crédito, promete uma melhora sobre os modelos mais antigos de pontuação de crédito com uso de técnicas lineares de estatística. Com a utilização de redes neurais consegue-se poder explicativo adicional, haja vista as complexas correlações e interações entre as variáveis independentes, que muitas vezes são realmente não-lineares.

Alguns autores são mais céticos quanto ao uso das redes neurais. Faraggi e Simon (1995) justificam que o desenvolvimento das redes neurais tem sido conduzido em grande parte por pesquisadores não estatísticos. Além disso, os autores ressaltam ainda que não há convicção sobre em que situações a utilização de redes neurais é preferível à utilização das técnicas estatísticas multivariadas.

Segundo Schumacher *et al.* (1996), apenas aplicações bem sucedidas das redes neurais são relatadas, e poucos pesquisadores têm demonstrado interesse na comparação entre redes neurais e técnicas estatísticas, no que se refere à eficiência. De acordo com Guimarães e Chaves Neto (2002), a discussão quanto à utilização e eficiência das redes neurais indica que o assunto é uma área ainda aberta a pesquisas, tanto no campo teórico como no campo das aplicações.

2.2.4 Precisão dos Modelos na Predição

Os testes estatísticos para medir a significância do modelo não garantem seu poder de predição. Para determinar a eficácia da predição, é necessário construir matrizes de classificação. Essas matrizes são construídas pela tabulação cruzada dos membros do grupo real com os do grupo previsto. Números na diagonal principal representam classificações corretas, e números fora da diagonal representam classificações incorretas (HAIR *et al.*, 2005).

Antes de obter a matriz de classificação, porém, é fundamental determinar o escore de corte. O ponto de corte é o limite em relação ao qual o escore de cada objeto é comparado para definir em qual grupo o objeto deve ser classificado (HAIR *et al.*, 2005).

Santos (2000) destaca que depois de determinada a pontuação de um potencial cliente, a decisão quanto à aprovação é da instituição. Como regra básica, pontuações de propostas de crédito abaixo do ponto de corte levam a decisão de recusa do crédito, enquanto pontuações acima do ponto de corte recomendam a aprovação.

Em termos práticos, Santos (2000) afirma que a determinação do ponto de corte deve considerar as políticas de crédito da empresa para definir as margens esperadas de contribuição de cada grupo (ganho com os bons e perda com os maus). Busca-se identificar o ponto de equilíbrio em que a margem de contribuição é igual a zero, ou seja, o ponto em que a perda com os maus clientes é compensada pelo ganho com os bons.

Na Figura 11 é apresentado um exemplo de pontuações que vão de 0 a 330, onde se determinou um ponto de corte de 210. A decisão para tomadores com escore inferior a esse valor é a recusa do crédito e com pontuação superior, a concessão do crédito. A distribuição hipotética de contas definidas como mau risco é apresentada na curva A, e as contas classificadas como bom risco na curva B. Com esse ponto de corte estabelecido, rejeitam-se 90% dos maus e também 20% dos bons clientes.

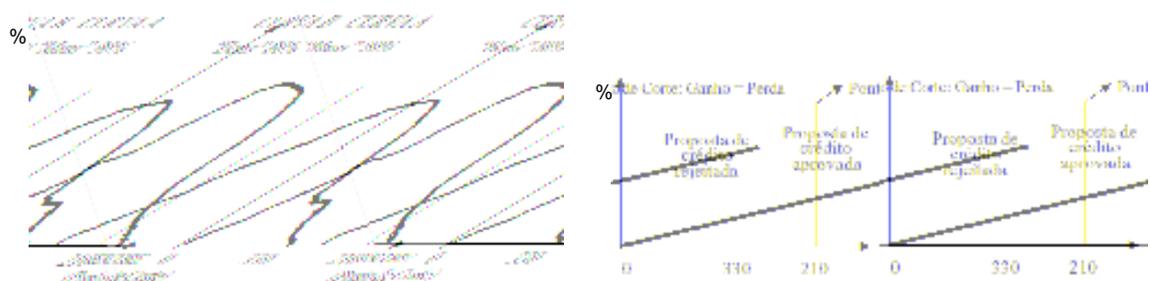


Figura 11 Exemplificação de um ponto de corte

Fonte: Santos (2000, p. 108)

Segundo Kimura *et al.* (2005), é importante avaliar mais detalhadamente os resultados obtidos pelo modelo, não somente os acertos como também os erros. Com altos percentuais de erros de classificação, a empresa pode vir a perder clientes que são bons pagadores, podendo diminuir sua participação no mercado. O autor destaca ainda que, muitas vezes, o impacto de negar crédito a um bom pagador pode conduzir a uma perda quase irreversível do cliente para a concorrência.

Da mesma forma, há também as classificações incorretas dos clientes que são inadimplentes. Ao sugerir que os clientes que não pagam são bons pagadores, a empresa pode incorrer em maiores perdas por inadimplência. Porém, dependendo da estratégia da empresa, estas perdas podem ser compensadas por um maior *market share*, resultando uma maior agressividade na concessão de crédito em comparação com a concorrência. Portanto, o gestor deve avaliar, de acordo com a estratégia da empresa, a melhor política de crédito. O modelo de previsão de crédito apenas possibilita uma indicação do perfil de pagamento dos clientes, mas não substitui o discernimento do tomador de decisão (KIMURA *et al.*, 2005).

Nos casos em que os tomadores obtenham pontuações próximas ao ponto de corte, Santos (2000) sugere que a decisão final da concessão do crédito aconteça aliando-se uma análise subjetiva do analista, como é apresentado na Figura 12.

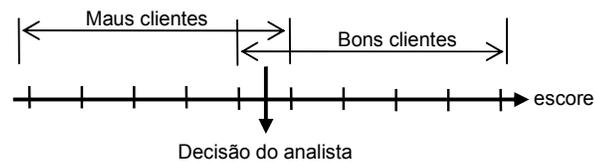


Figura 12 Análise objetiva e subjetiva aliadas na concessão de crédito

Segundo Hair *et al.* (2005), o ponto de corte ótimo será modificado dependendo se os tamanhos dos grupos são iguais ou diferentes. Se os grupos são de tamanhos iguais, o escore de corte ótimo será o ponto médio entre os centróides (médias) dos dois grupos. Se os grupos são de tamanhos diferentes, mas são representativos das proporções da população, utiliza-se uma média ponderada dos centróides dos grupos, calculada segundo a Equação 03.

$$Z_{CU} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B} \quad (03)$$

onde,

Z_{CU} = ponto de corte ponderado para grupos com tamanhos diferentes

N_A = número de observações do grupo A

N_B = número de observações do grupo B

Z_A = centróide para o grupo A

Z_B = centróide para o grupo B

Na Figura 13 é ilustrado o conceito de ponto de corte para grupos de tamanhos diferentes. Observa-se que se o grupo A é muito menor que o grupo B e, se o ponto de corte não-padronizado fosse utilizado, nenhum dos indivíduos no grupo A seria mal classificado, mas uma parte considerável dos que estão no grupo B seria mal classificada.

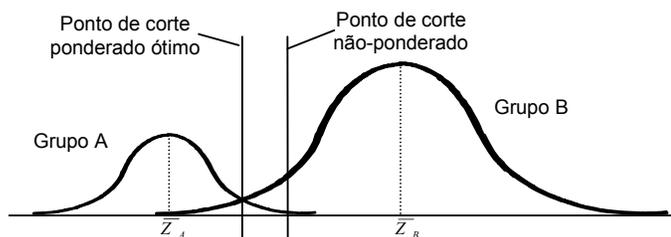


Figura 13 Ponto de corte com tamanhos diferentes de amostras

Fonte: Hair *et al.* (2005, p. 225)

Um critério bastante utilizado na prática (PICININI *et al.*, 2003) para avaliar a precisão dos modelos construídos é o teste de Kolmogorov-Smirnov (KS). O teste KS é uma técnica não paramétrica utilizada para verificar se duas amostras foram extraídas de uma mesma população. Este teste se baseia na distribuição acumulada dos escores dos clientes considerados como bons e maus.

Como na construção de modelos de previsão de risco as amostras são grandes, a tendência é que sempre se rejeite a hipótese de igualdade nas distribuições. Portanto, deve ser considerado como melhor modelo aquele que possuir o maior valor no teste, indicando uma maior separação entre bons e maus. Quanto à definição de um valor ideal para o teste KS, Picinini *et al.* (2003, p. 465) sugerem: “o teste de Kolmogorov-Smirnov (KS) é utilizado no mercado financeiro como um dos indicadores de eficiência de modelos de *credit scoring*, sendo que o mercado considera um bom modelo aquele que apresente um valor de KS igual ou superior a 30”.

Kassai e Kassai (1998) destacam que uma vez obtido um grau de precisão aceitável, o modelo está aprovado. Assim, enquanto um nível desejado para predição do modelo não é alcançado, seria necessário buscar novos dados e continuar o processo de avaliação do modelo. Contudo o autor não sugere qual seria esse grau de precisão aceitável. Segundo Picinini *et al.* (2003, p. 465): “modelos de *credit scoring* com taxas de acerto acima de 65% são considerados bons por especialistas”.

Na tentativa de responder essa questão, Hair *et al.* (2005) sugerem determinar qual seria o percentual que poderia ser classificado corretamente ao

acaso (por chances), ou seja, sem a utilização do modelo de pontuação. Quando as amostras são de mesmo tamanho esse percentual é de 50%, porém, nos casos em que os tamanhos dos grupos são desiguais, a determinação desse percentual de classificação é um pouco mais complicada. Trata-se do critério de chances proporcionais que é obtido segundo a Equação 04.

$$C_{PRO} = p^2 + (1 - p)^2 \quad (04)$$

onde,

p = proporção de indivíduos do grupo 1;

$1-p$ = proporção de indivíduos do grupo 2

A questão que se coloca agora é: quanto maior deve ser a classificação de sucesso do modelo em comparação com a proporção de classificação ao acaso. Segundo Hair *et al.* (2005), a precisão de classificação deve ser pelo menos um quarto maior do que a obtida ao acaso. Por exemplo, se as amostras são de tamanhos iguais e, portanto, tem-se 50% de classificação correta por chances, a precisão de classificação deverá ser pelo menos de 62,5% (= 1,25 x 50%).

Analisar as observações que foram mal classificadas é um ponto relevante para o entendimento do relacionamento das variáveis independentes na predição. Porém, ao examinar esses casos com as variáveis independentes, é possível identificar tendências não-lineares ou outras relações ou atributos que conduzam à má classificação (HAIR *et al.*, 2005).

Nesse sentido, Saunders (2000) sugere que o uso de uma rede neural possivelmente possa reduzir os erros de classificação do tipo 1 e 2 de uma base histórica de empréstimos para bem próximo de zero. O autor classifica como erro do tipo 1 a classificação incorreta de um cliente ruim como sendo bom, e o erro do tipo 2 como a classificação incorreta de um bom cliente como ruim.

2.3 APLICAÇÕES DOS MODELOS DE PREVISÃO

Modelos de previsão de risco de crédito são encontrados em praticamente todos os tipos de análise de crédito, desde crédito ao consumidor até empréstimos comerciais. A idéia destes sistemas está na identificação de fatores-chave que

possam determinar a probabilidade de inadimplência, obtendo, através da sua ponderação, uma pontuação quantitativa (SAUNDERS, 2000). Nessa seção são apresentados alguns estudos existentes na bibliografia a respeito do desenvolvimento de modelos de previsão de risco de crédito para pessoas físicas e jurídicas, utilizando as técnicas discutidas: análise discriminante, regressão logística e redes neurais.

Altman *et al.* (1979) desenvolveram um modelo utilizando análise discriminante com o objetivo de alertar os credores na concessão de financiamento às empresas brasileiras que apresentavam tendências de se tornarem insolventes ou incapazes de amortizar os empréstimos na data prevista. Utilizando uma amostra de 58 empresas, os autores conseguiram acertos de 88% para as empresas consideradas boas e 87% para as empresas consideradas ruins.

Almeida (1995) apresentou em seu artigo uma visão do potencial e do funcionamento do uso de redes neurais em administração. As redes neurais estão provocando um grande impacto em algumas empresas americanas permitindo-lhes obter reduções consideráveis de custos em análises de crédito. Conclui que as redes neurais têm mostrado resultados positivos em um espaço de tempo mais curto que outras técnicas de criação de bases de conhecimentos como os sistemas especialistas, e também se mostraram bem mais flexíveis que uma técnica estatística. Utilizando uma base de 45 empresas inadimplentes e 135 adimplentes, as redes neurais foram capazes de classificar corretamente aproximadamente 70% das empresas.

Desai *et al.* (1996) compararam redes neurais, análise discriminante linear e regressão logística para a construção de modelos de escore de crédito. Os autores avaliaram o percentual de acerto total dos modelos e neste aspecto o modelo neural foi bastante superior aos demais. Porém ao analisar o percentual de clientes bons e maus corretamente classificados o modelo logístico teve desempenho equivalente ao neural.

Almeida e Dumontier (1996) publicaram um trabalho em que propuseram uma abordagem estruturada de exploração de redes neurais para avaliação dos riscos de inadimplência do setor de transporte de carga rodoviário francês. A técnica

de redes neurais foi comparada à regressão logística, e os resultados mostraram que o desempenho da primeira não foi significativamente superior aos obtidos com a utilização da regressão logística. Porém, os autores ressaltaram que estudos futuros poderiam aprofundar a compreensão do processo de falência através do uso de redes neurais, possibilitando melhores resultados em outros setores de atividades.

Almeida e Siqueira (1997) exploraram o uso de redes neurais na previsão de falência de bancos brasileiros. Os resultados de classificação das redes neurais foram comparados com a técnica de regressão logística. Apesar das redes neurais não apresentarem resultados superiores, 67% de classificação correta contra 68% da regressão logística, um fator diferencial a ser destacado foi a capacidade de considerar bancos que foram desprezados pela regressão logística devido à falta de dados.

Arminger *et al.* (1997) fizeram uma comparação entre três técnicas multivariadas para prever se um cliente pagará ou não um empréstimo. As variáveis independentes para construção dos modelos foram sexo, tempo de serviço, idade, ter carro, ter telefone e estado civil. Os resultados demonstraram acertos na classificação dos clientes de 67,6% para a regressão logística, 66,4% para a árvore de decisão e 65,2 para as redes neurais. Num esforço adicional os autores buscaram uma previsão combinada utilizando as três técnicas, porém o resultado não foi satisfatório, sendo pior que o obtido pela regressão logística.

Steiner *et al.* (1999) utilizaram sistemas especialistas probabilísticos e redes neurais, ambas da área da Inteligência Artificial, na análise de crédito bancário para pessoa física, com o objetivo de prever o comportamento de futuros clientes quanto à inadimplência. Tendo por base dados históricos de 2.855 clientes de um banco alemão, foram obtidos padrões, mostrando, para diagnósticos futuros, em qual situação é interessante para o banco fazer a concessão do crédito.

Züge e Chaves Neto (1999), utilizando a análise discriminante, desenvolveram uma regra de reconhecimento de padrões para classificar empresas, levando em consideração diversos índices econômico-financeiros, resultantes da análise dos demonstrativos contábeis. Foram analisadas 60 empresas tomadoras de crédito junto ao BRDE (Banco Regional de Desenvolvimento do Extremo Sul). O

desempenho do modelo obtido se revelou satisfatório, sendo que o total de classificações corretas ficou em 93,3%. Os autores concluíram ainda que a utilização do modelo aliado à análise de outros fatores, como conjuntura econômica, qualidade de gestão administrativa das empresas, entre outros, melhora sensivelmente o processo, levando a tomada de decisões de crédito mais embasadas e acertadas.

Selau (2000) apresentou um estudo de redes neurais destacando sua importância como ferramenta para o reconhecimento de padrões e sua superioridade em comparação com outras técnicas. Em uma aplicação na área de crédito a autora utilizou uma base de 1000 créditos concedidos em um banco alemão, sendo 700 bons pagadores e 300 maus pagadores. Obteve resultados superiores com a aplicação de redes neurais para a previsão do risco de crédito quando comparada com as técnicas de regressão logística e árvore de decisão.

Saunders (2000) apresentou e comentou o modelo de pontuação Z de Altman, que é um modelo classificatório para empresas tomadoras de crédito, obtido através do método de análise discriminante. O primeiro comentário do autor foi quanto à linearidade do modelo, já que as variáveis em estudo mostram que o processo de falência de uma empresa tem um comportamento altamente não-linear. A segunda questão foi sobre o fato de o modelo ser construído essencialmente baseado em coeficientes contábeis, dados esses que aparecem apenas em intervalos regulares e que de forma geral se baseiam em princípios de contabilidade históricos ou de registro. Por último, questionou sobre a eficiência do modelo em detectar uma empresa cuja condição esteja em rápida deterioração.

Santos (2000) detalhou o modelo desenvolvido em 1978 pelo Professor Kanitz da Faculdade de Economia e Administração da Universidade de São Paulo, o primeiro estudo realizado no Brasil com a finalidade de prever falências. Através da análise discriminante e com base em uma amostra de 30 empresas, sendo 15 falidas e 15 sem problemas financeiros, encontrou um modelo com percentual de acerto de 80% para empresas solventes e de 68% para empresas insolventes.

Marques e Lima (2002) analisaram o desempenho de empresas para fins de concessão de crédito através de variáveis contábeis. Foram comparados os resultados dos modelos de previsão com uso de análise discriminante e regressão

logística, obtendo respectivamente os percentuais de acerto de 91,67% e 93,18% na classificação das empresas. Os autores afirmaram ainda que foi de suma importância o descarte de *outliers* e variáveis que estariam deturpando a separação dos perfis.

Scarpel e Milioni (2002) utilizaram um modelo econométrico do tipo *logit* em conjunto com o modelo de programação linear inteira, com o objetivo de auxiliar a análise de crédito utilizando a estimativa da probabilidade de solvência de empresas, minimizando a soma dos custos de oportunidade e inadimplência. Os autores concluíram que a utilização conjunta desses modelos possibilitou eliminar as limitações encontradas quando eles são utilizados isoladamente nas decisões de concessão de crédito. Foi encontrada 87% de eficiência global, já que o modelo previu corretamente a insolvência de todas as 7 empresas efetivamente insolventes e falhou prevendo a insolvência de 5 das 32 empresas que na verdade eram solventes.

Assaf e Brito (2005) desenvolveram um modelo de previsão de risco de crédito para grandes empresas, utilizando regressão logística. O modelo foi construído com base em uma amostra de empresas de capital aberto classificadas como solventes e insolventes no período entre 1994 e 2004, e informações de índices contábeis, utilizados para caracterizar a situação econômico-financeira das empresas. Os resultados alcançados mostraram que o modelo proposto alcançou a proporção de 88,3% de classificações corretas, evidenciando que as demonstrações contábeis fornecem informações que permitem uma boa classificação das empresas.

Kimura *et al.* (2005), através do uso do procedimento de redes neurais, propuseram um modelo de previsão de risco de crédito com base em uma amostra aleatória de 2475 clientes de uma importante rede de lojas brasileira. A melhor rede propiciou 79% de acertos sobre o perfil de pagamento. Os autores destacaram ainda que, tendo sido considerado na pesquisa um número reduzido de variáveis de cadastro, os resultados sugerem que as redes neurais podem representar uma promissora técnica para análise de concessão de crédito ao consumidor.

Gouvêa e Gonçalves (2006), a partir de uma amostra de dados, fornecida por uma grande instituição financeira brasileira, desenvolveram dois modelos,

aplicando-se duas técnicas para a classificação de clientes: redes neurais e algoritmos genéticos. Os resultados obtidos pelos modelos de rede neural e algoritmo genético foram satisfatórios, sendo o primeiro modelo superior com percentual de acerto de 69,4% contra 67,6%. Os autores utilizaram ainda o teste KS para verificar a eficiência dos modelos, obtendo 39 para a técnica de redes neurais e 34 para os algoritmos genéticos.

2.4 CONCLUSÃO

Este capítulo apresentou o referencial teórico que alicerça a problemática da construção dos modelos de previsão de risco de crédito bem como a descrição das três técnicas multivariadas que serão utilizadas neste trabalho. A análise de diversos trabalhos publicados na área evidencia a importância do assunto para o ambiente de decisões financeiras e sua possibilidade de crescimento. Percebe-se também que não há um detalhamento no método de construção dos modelos de previsão de risco de crédito. Neste sentido, o método apresentado na sequência se propõe descrever em detalhes as etapas para a construção de tais modelos, desde definições a respeito da população até passos para a implantação no sistema da empresa.

3 SISTEMÁTICA PARA DESENVOLVIMENTO DO MODELO PRC

Neste capítulo são descritas as etapas e as características para o desenvolvimento do Modelo de Previsão de Risco de Crédito (Modelo PRC). A metodologia proposta, constituída de seis etapas (ver Figura 14), foi desenvolvida a partir dos tópicos abordados na revisão da literatura. As seis etapas são explicadas na seqüência, contemplando desde os primeiros detalhes para iniciar a pesquisa até os passos para a implantação do modelo escolhido.

Delimitação da População	<ul style="list-style-type: none">• existência de histórico de crédito;• seleção da população alvo para a qual se dirige o modelo PRC;• definição de desempenho satisfatório e insatisfatório.
Seleção da Amostra	<ul style="list-style-type: none">• identificação de variáveis disponíveis no sistema da empresa;• definição do período e tamanho da amostra;• validação da consistência e preenchimento dos dados;• separação da amostra para análise e teste.
Análise Preliminar	<ul style="list-style-type: none">• escolha de variáveis para entrar na modelagem;• agrupamento de atributos de variáveis;• criação das variáveis <i>dummies</i>.
Construção do Modelo	<ul style="list-style-type: none">• escolha da técnica estatística multivariada;• determinação do <i>software</i> a ser utilizado;• seleção de variáveis independentes;• verificação de suposições das técnicas.
Escolha do Modelo	<ul style="list-style-type: none">• percentual de classificações corretas;• valor do teste KS para duas amostras.
Passos para Implantação	<ul style="list-style-type: none">• preparação do sistema de informação;• determinação do ponto de corte;• indicações e cuidados na utilização do modelo;• relatórios de revisão e acompanhamento do modelo.

Figura 14 Etapas para desenvolvimento do Modelo PRC

3.1 DELIMITAÇÃO DA POPULAÇÃO

A primeira etapa da sistemática proposta contempla a delimitação da população e compreende os passos: (i) existência de histórico de crédito; (ii) seleção da população alvo para a qual se dirige o modelo PRC; (iii) definição de desempenho satisfatório e insatisfatório.

A suposição básica para construir um modelo de previsão de crédito é que os clientes têm o mesmo padrão de comportamento ao longo do tempo (GOUVÊA; GONÇALVES, 2006). Portanto, considerando que a construção do modelo é exclusivamente baseada na experiência do uso do crédito pela empresa, todos os dados utilizados no desenvolvimento são oriundos dos registros deste negócio. Os dados da amostra têm que constituir toda a informação conhecida dos clientes na hora da concessão do crédito e também seus *status* subseqüentes como bons e maus pagadores.

Antes da definição dos parâmetros para a seleção da amostra, é necessário decidir para qual segmento da população o modelo construído vai ser utilizado. Em empresas de pequeno e médio porte, onde há somente um tipo de crédito, pode ser a população (todos os clientes). Já para empresas grandes, onde são oferecidos diversos produtos de crédito, a população para o estudo deve ser limitada por tipo de produto.

Para a construção do modelo, é imprescindível que existam créditos concedidos pelo negócio e que os resultados da concessão tenham sido avaliados. Portanto, para que se possa desenvolver o modelo, inicialmente deve-se definir os conceitos de desempenho aceitável e inaceitável. Conforme o modelo proposto neste capítulo, a avaliação do desempenho de crédito é feita segundo uma escala dicotômica, definindo o cliente como bom ou mau pagador.

Para Gouvêa e Gonçalves (2006), além dos clientes bons e maus, também existem os clientes que devem ser excluídos da análise, como por exemplo, os clientes que trabalham na empresa e os clientes indeterminados, que estão na fronteira entre aqueles que seriam considerados bons ou maus, sem uma posição quanto aos pagamentos. Na prática, ao construir o modelo, as empresas costumam considerar apenas os clientes bons e maus, devido à facilidade de trabalhar com modelos de resposta dicotômica.

Existem, portanto, 4 grupos que devem ser separados no total de créditos concedidos: (i) os clientes que nunca utilizaram o crédito – sem uso; (ii) os clientes com pouco ou nenhum atraso – bons; (iii) os clientes em faixas de atrasos intermediárias – indeterminados; (iv) os clientes com atrasos consideráveis – maus.

A definição de atrasos consideráveis deve ser feita pelo concessor, que definirá os atrasos que podem ser aceitos pelo negócio. Na construção do modelo, somente são utilizados os grupos de clientes bons e maus para facilitar a análise e para acentuar a separação de perfis.

3.2 SELEÇÃO DA AMOSTRA

Na etapa de seleção da amostra são definidos os seguintes critérios: (i) identificação de variáveis disponíveis no sistema da empresa; (ii) definição do período e tamanho da amostra; (iii) validação da consistência e preenchimento dos dados; (iv) separação da amostra para análise e teste.

A análise das informações constantes no banco de dados da empresa pode ser feita pela observação detalhada da proposta de crédito. Através dela é possível listar as possíveis variáveis que poderão fazer parte do modelo final e, portanto, devem ser listadas na amostra de estudo. Dentre as possíveis informações selecionadas, chamadas também de variáveis demográficas, pode-se citar: sexo, idade, escolaridade, estado civil, tipo de ocupação, tipo de residência, tempo no emprego atual, entre outras (MESTER, 1997).

Lewis (1992) comenta a respeito dos tipos de variáveis independentes utilizadas em estudos em diversos países. De forma geral, as variáveis demográficas são comuns em todos os modelos de previsão de risco de crédito, incluindo também em muitos casos características referentes à operação de crédito em análise, como valor e número de parcelas. Nos Estados Unidos é comum a utilização de informações positivas sobre os clientes, como tempo desde o último empréstimo e valor de crédito utilizado. O autor sugere ainda que seja avaliada a inclusão de informações na proposta de crédito consideradas importantes para uma avaliação futura do modelo construído.

Para definição do período para extração da amostra é necessário observar um tempo entre a concessão do crédito e a verificação de seu desempenho de pagamento. Lawrence (1992) sugere que os clientes que farão parte da amostra tenham sido incluídos há no mínimo 12 a 18 meses. Este tempo decorrido é usualmente suficiente para que se verifique a ocorrência de pagamentos ou de

parcelas em aberto com muitos dias de atraso, definindo o cliente mau, e também de se consolidar o comportamento de pagamento do cliente bom.

Na utilização de técnicas multivariadas, o tamanho da amostra depende do número de variáveis independentes que farão parte do estudo para construção do modelo final e, principalmente, se a amostra contempla adequadamente a maioria das combinações que podem ser geradas a partir dessas variáveis. Hair *et al.* (2005) sugerem a utilização de uma proporção de 20 observações para cada variável independente.

No que se refere à construção de modelos de previsão de risco de crédito, não existe na literatura definição de tamanho adequado para a amostra a ser utilizada. Lewis (1992) propõe a busca de 1500 clientes bons e 1500 maus, justificando que este número é grande o suficiente para permitir a construção de sistema de previsão e, ao mesmo tempo, evitar despesas desnecessárias na coleta da amostra. Porém, o tamanho da amostra na maioria dos casos não é um problema, pois quando se trata de empresas que concedem crédito à pessoa física, há abundância de dados históricos (VASCONCELLOS, 2004).

De posse do banco de dados, é efetuada uma análise exploratória, onde todos os campos são analisados quanto ao seu conteúdo. Neste momento, devem ser verificadas questões quanto à qualidade de preenchimento, consistência dos campos e presença de observações faltantes (*missing*), eliminando dados inconsistentes ou atípicos.

Um último ponto a ser considerado quanto à amostra é a questão da divisão entre amostra de análise e teste, a fim de evitar qualquer tipo de tendenciosidade. De acordo com Assaf Neto e Brito (2005), ao se testar o modelo com a mesma amostra utilizada para sua construção, pode-se concluir que o seu desempenho é bom quando, na verdade, ele pode funcionar bem apenas para estas observações. Portanto, para verificar se o poder preditivo do modelo é mantido para outras amostras provenientes da mesma população, são necessários testes para a sua validação.

Não existem regras fixas quanto à partição da amostra. Hair *et al.* (2005) argumentam que a escolha quanto à divisão da amostra deve ficar por conta do pesquisador. Devido à importância que a construção do modelo tem em relação ao seu teste, propõe-se a divisão da amostra total conforme sugere Haykin (2001): 80% da amostra total para análise e 20% para teste do modelo final. A separação das amostras é feita através de rotina computacional, gerando-se uma variável aleatória uniformemente distribuída utilizada para alocar, ao acaso, os casos às respectivas amostras.

3.3 ANÁLISE PRELIMINAR

A etapa de análise preliminar contempla os passos: (i) escolha de variáveis para entrar na modelagem; (ii) agrupamento de atributos de variáveis; (iii) criação das variáveis *dummies*.

O primeiro passo, antes de começar as análises das informações do banco de dados, trata-se da escolha das variáveis que entrarão na análise, podendo vir a integrar o modelo final. Segundo Smith (1996), não existe uma metodologia consolidada para esse processo de escolha, dependendo de testes estatísticos e do bom senso de cada pesquisador.

Através da análise da associação entre as variáveis independentes e a variável dependente (tipo de cliente: bom ou mau) é possível selecionar os atributos que poderão entrar na fase seguinte (construção do modelo). Atributos são as possíveis respostas para cada uma das variáveis. Por exemplo, feminino e masculino são os atributos da variável sexo do cliente.

Com uso de tabelas de contingência, calcula-se o risco relativo (RR) para os atributos das variáveis independentes, dividindo-se o percentual de bons clientes pelo percentual de maus de cada atributo. Quanto mais os percentuais de bons e maus diferirem para os atributos de uma mesma variável, maior será a utilidade dessa variável para o prognóstico de desempenho futuro (LEWIS, 1992). Por exemplo, se a mesma fração de bons e maus clientes têm casa própria ou alugada, essa variável não provê nenhuma informação que ajude a estabelecer a probabilidade de um cliente vir a se tornar bom ou mau pagador.

Para variáveis com muitos atributos possíveis, como por exemplo, CEP e profissão, são necessários agrupamentos. Isso é feito com o objetivo de diminuir o número de atributos a serem analisados e, principalmente, para obter representatividade nos agrupamentos, de forma a aumentar a chance de fazerem parte do modelo final. A medida utilizada para fazer estes agrupamentos é também o resultado do risco relativo.

Como regra geral, os atributos são divididos, segundo o valor do risco relativo, nas seguintes classes: péssimo (RR < 0,50); muito mau (RR entre 0,50 e 0,67); mau (RR entre 0,67 e 0,90); neutro (RR entre 0,90 e 1,10); bom (RR entre 1,10 e 1,50); muito bom (RR entre 1,50 e 2,00) e excelente (RR maior que 2,00). Atributos classificados como neutros não são utilizados na análise, já que não têm diferença forte entre os grupos bom e mau (LEWIS, 1992).

Após a seleção dos atributos que farão parte da análise multivariada, passa-se para a criação de uma variável *dummy* para cada um deles (ex.: cada nível de escolaridade será uma variável *dummy*). Essa variável só assume dois valores (0 ou 1). Com esse artifício evitam-se problemas decorrentes da não linearidade dos atributos no cálculo da análise multivariada.

3.4 CONSTRUÇÃO DO MODELO

A construção do modelo compreende os passos: (i) escolha da técnica estatística multivariada; (ii) determinação do *software* a ser utilizado; (iii) seleção de variáveis independentes; (iv) verificação de suposições das técnicas.

Uma vez reduzidos os dados a agrupamentos de atributos, cuidadosamente escolhidos para todas as características, e criação das respectivas variáveis *dummies*, cada analista escolhe o método a ser utilizado para a modelagem. Neste estudo sugere-se a utilização de análise discriminante, regressão logística e redes neurais. Tais métodos encontram-se entre os mais utilizados para a construção de tais modelos e por isso foram os escolhidos (DESAI *et al.*, 1996; ALMEIDA; DUMONTIER, 1996; ARMINGER *et al.*, 1997; MARQUES; LIMA, 2002).

A análise discriminante e regressão logística são técnicas estatísticas multivariadas que adotam abordagens diferentes, existindo a possibilidade de uma dessas técnicas ter sucesso quando a outra falha, principalmente no que diz respeito ao cumprimento de alguns pressupostos. As redes neurais também fazem parte da proposta em função de sua habilidade de lidar com efeitos não lineares e descontínuos, identificando relações que os métodos estatísticos usuais não contemplam.

A escolha do *software* a ser utilizado é um passo importante, devendo-se verificar suas características quanto às necessidades das análises e sua facilidade de uso. Para a construção dos modelos com uso das técnicas de análise discriminante e regressão logística, sugere-se a utilização de um pacote estatístico comercial. Para o treinamento e teste das redes neurais, indica-se a busca de um módulo adicional aos pacotes estatísticos convencionais ou um *software* específico para o desenvolvimento da técnica.

Nas técnicas multivariadas recomendadas para a modelagem, há sempre uma única variável dependente (tipo de cliente: bom ou mau), que é também transformada em variável *dummy*, sendo 0 – mau e 1 – bom. Há também diversas variáveis demográficas, como por exemplo, sexo e idade, que serão consideradas as variáveis independentes para a construção dos modelos.

A construção de um modelo adequado, com a utilização de qualquer uma das técnicas propostas, é uma tarefa complexa. É necessária, por exemplo, a avaliação de variáveis que devem entrar ou sair da análise para evitar problemas de multicolinearidade. Muitas vezes, devido à presença de multicolinearidade (correlação entre as variáveis independentes), podem ocorrer trocas de sinais dos pesos no modelo final.

Neste sentido, Smith (1996) sugere que a inclusão das variáveis independentes no modelo aconteça conforme sua associação com a variável dependente. Num segundo momento, se o modelo ainda não atingir um nível de desempenho satisfatório, passa-se para a inclusão das variáveis com menor grau de explicação. O autor comenta, ainda, que esse procedimento é válido tanto para métodos estatísticos como para as redes neurais. Para a seleção inicial das variáveis

independentes que devem compor o modelo, pode-se utilizar o método *stepwise*, incorporado em muitos pacotes estatísticos, que automaticamente seleciona a melhor combinação de variáveis independentes para entrada no modelo.

Para seguir com a avaliação e utilização dos modelos construídos é necessária a observação dos pressupostos para utilização das técnicas estatísticas. A metodologia de redes neurais é bem mais flexível que as demais técnicas estatísticas, sendo que nenhum pressuposto precisa ser verificado. As redes neurais não pressupõem um modelo ao qual os dados devem ser ajustados, já que o modelo é gerado pelo processo de aprendizagem (CORRAR *et al.*, 2007).

Os principais pressupostos da análise discriminante a serem verificados são: normalidade multivariada, homogeneidade de matrizes de variância e ausência de multicolinearidade. Para avaliar a normalidade multivariada utiliza-se o teste de Kolmogorov-Smirnov (KS) para uma amostra e para testar a homogeneidade de variâncias, o teste denominado Box's M que testa a hipótese nula de igualdade de matrizes de variância. A ausência de multicolinearidade é atendida com o uso do método *stepwise* para identificação de variáveis independentes para compor o modelo final.

Na regressão logística, o único pressuposto a ser verificado é o da ausência de multicolinearidade. Assim como na análise discriminante, tal suposição é atendida com a utilização do método *stepwise* para a seleção das variáveis independentes.

3.5 ESCOLHA DO MODELO

Para escolha do melhor modelo, utilizando cada uma das três técnicas recomendadas, podem-se avaliar duas medidas de desempenho: (i) percentual de classificações corretas; (ii) o valor do teste KS para duas amostras.

Primeiramente, para utilização das duas medidas, é necessário definir o procedimento de separação das classificações de bons e maus. Como o resultado final obtido do modelo, com utilização de qualquer uma das técnicas, é uma escala de 0 a 1 (a pontuação baseada no risco de cada cliente), define-se como mau o

cliente com resposta menor que 0,5; caso contrário será classificado como um cliente bom.

O percentual de acerto nas classificações deve ser avaliado através do cruzamento dos resultados observados e previstos pelo modelo, conforme apresentado na Figura 15. Na diagonal principal ficam os casos classificados corretamente, clientes maus que foram previstos como maus e clientes bons previstos como bons. Desta forma, a taxa de acerto é medida pela divisão da quantidade de clientes corretamente classificados pelo total de clientes que fizeram parte da análise. Especialistas consideram satisfatórios os modelos com taxa de acerto superior a 65% (PICININI *et al.*, 2003).

Previsto \ Observado	BOM	MAU	TOTAL
BOM	Acertos na classificação dos bons	Bons classificados como maus	Total de bons na amostra
MAU	Maus classificados como bons	Acertos na classificação dos maus	Total de maus na amostra
TOTAL	Total de bons previsto pelo modelo	Total de maus previsto pelo modelo	Total de clientes

Figura 15 Verificação de acerto nas classificações do modelo

O teste KS tem como característica a simplicidade. O que se busca é determinar a diferença máxima entre duas distribuições acumuladas. As duas subpopulações (bons e maus), traduzidas pelos seus respectivos resultados previstos pelo modelo, são dispostas em distribuição cumulativas de freqüências. Determinam-se as diferenças entre as distribuições de bons e maus para cada resultado previsto, sendo o valor do teste KS a maior dessas diferenças em módulo (ver Figura 16). Obtendo-se uma diferença maior que 30%, pode-se considerar que o modelo é eficiente na predição dos dois grupos (PICININI *et al.*, 2003). O valor pode ainda ser comparado com um valor crítico a uma significância de 1% conforme Equação 05.

$$KS_{1\%} = 1,63 \sqrt{\frac{n_1 + n_2}{n_1 \times n_2}} \quad (05)$$

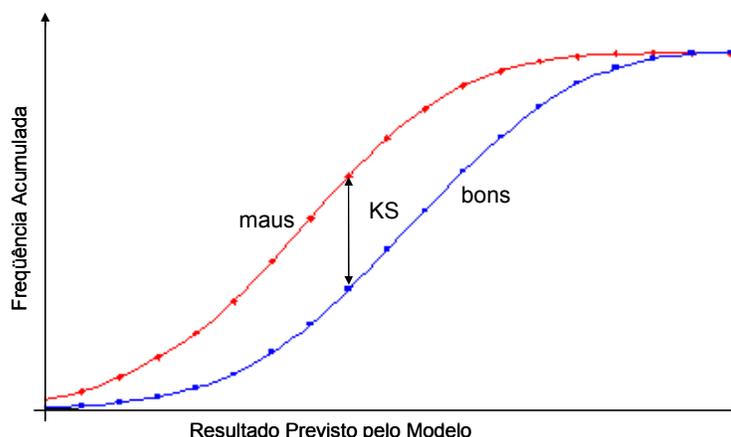


Figura 16 Exemplificação do método de obtenção do valor do teste KS

3.6 PASSOS PARA IMPLANTAÇÃO

Nesta última etapa são sugeridos os passos para a implantação do modelo escolhido na empresa: (i) preparação do sistema de informação; (ii) determinação do ponto de corte; (iii) indicações e cuidados na utilização do modelo; (iv) relatórios de revisão e acompanhamento do modelo.

Por meio dos critérios previamente definidos, o melhor modelo é escolhido. Com isso deve-se programar a implantação do modelo junto à empresa. O Departamento de Tecnologia da Informação da empresa deve adequar seus sistemas para receber o modelo final e programar sua utilização junto às demais áreas envolvidas (GOUVÊA; GONÇALVES, 2006).

Quanto à definição do ponto de corte, Vasconcellos (2004) propõe um procedimento baseado na taxa de inadimplência. Primeiramente, ordenam-se os clientes segundo seu nível de risco (score obtido pelo modelo) e calcula-se o resultado esperado de inadimplência para cada nível. Espera-se uma diminuição da inadimplência conforme aumenta a pontuação. Segundo o autor, o processo de definição do ponto de corte tem por objetivo buscar o ponto onde o risco se torna tão alto que a inadimplência se torna não sustentável.

Além de identificar o comportamento da inadimplência, a pontuação obtida com o modelo pode também ser utilizada para definir formas de gerenciar a concessão de crédito, como exemplificado por Queiroz (2006). Como pode ser observado na Figura 17, a autora sugere algumas políticas de crédito (não aprovar, realizar análise de todos os documentos, análise de documentos de renda, abordar clientes sem análise de documentos e clientes VIPs) a partir das faixas de escore.

Uma das questões mais importantes na implantação do modelo é garantir que as propostas de crédito sejam avaliadas nas mesmas condições em que foram avaliados os clientes que constituíram a amostra utilizada no desenvolvimento do modelo. Deve-se garantir, portanto, que as informações obtidas no momento da análise de novas propostas apresentem condições semelhantes às coletadas no passado (VASCONCELLOS, 2004).

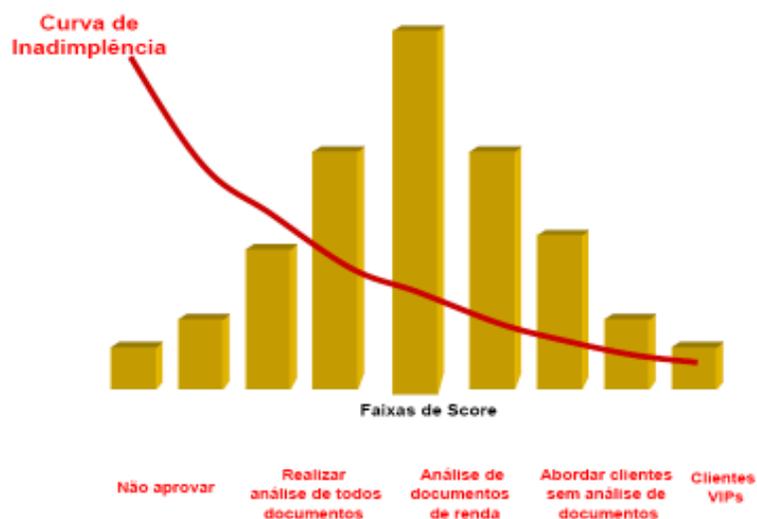


Figura 17 Exemplo de política de crédito segundo a inadimplência

Fonte: Queiroz (2006, p. 5)

Pereira *et al.* (2002) sugerem que, após um ano de utilização do modelo, uma revisão seja feita, seguindo os mesmos passos para a construção do modelo original. Depois disso, os autores afirmam que a revisão é obrigatória somente se houver mudança significativa na inadimplência, na lucratividade, nos prazos ou

condições do negócio e, principalmente, no perfil da população. Tais alterações devem ser monitoradas através de relatórios de acompanhamento para o modelo.

3.7 CONCLUSÃO

A sistemática apresentada neste capítulo descreveu a forma de construção de um modelo de previsão de risco de crédito, abordando a delimitação da população e os detalhes para a seleção da amostra com a escolha das variáveis. . Também foram apresentados os passos para as análises preliminares dos dados e a construção dos modelos com uso das três técnicas escolhidas: análise discriminante, regressão logística e redes neurais. Por fim, apresentaram-se os métodos para escolha dos melhores modelos e passos para implantação nos sistema da empresa. Na próxima seção são apresentados os resultados da aplicação desta sistemática.

4 APLICAÇÃO DO MODELO PRC

Nesta seção são apresentados os resultados obtidos em cada etapa da aplicação do modelo proposto na seção 3.

4.1 DELIMITAÇÃO DA POPULAÇÃO

No desenvolvimento desta pesquisa, foram utilizadas informações sobre os clientes de uma rede de farmácias com unidades em todo o Rio Grande do Sul. É oferecido aos clientes um cartão de crédito próprio como forma de facilitar o pagamento das compras. É importante salientar, ainda, que as informações aqui apresentadas foram parcialmente transformadas, de forma a preservar o sigilo dos dados.

Como a empresa estudada tem, atualmente, somente um produto de crédito, o alvo da pesquisa é, portanto, toda a população de clientes deste produto. A empresa oferece aos seus clientes um cartão próprio, utilizado para efetuar o pagamento das compras feitas nas lojas da rede. O pagamento pode, ainda, ser parcelado em até três vezes, com vencimento único de fatura enviada ao endereço do cliente, semelhante a um cartão de crédito convencional.

De acordo com a determinação da qualidade do crédito desejada pelo conessor, chega-se à definição dos grupos de clientes segundo os atrasos no período analisado. Para a empresa em estudo, o cliente bom é definido como aquele que tem atrasos de até 30 dias e os clientes maus são aqueles com pelo menos um atraso superior a 60 dias. Como indefinidos são classificados os clientes com atrasos entre 31 e 60 dias. No desenvolvimento dos modelos, este último grupo de clientes é excluído da amostra de forma a conseguir maior poder de discriminação. Além destes três grupos de clientes, foi separado um quarto grupo composto dos clientes que não tiveram nenhuma compra com o cartão no período em estudo. Este grupo de clientes é chamado sem uso.

4.2 SELEÇÃO DA AMOSTRA

A identificação das informações disponíveis no sistema da empresa, que serviram como variáveis independentes para a análise, foi feita a partir da proposta que é preenchida pelos clientes no momento da solicitação do crédito. Num primeiro momento, foram consideradas as 16 variáveis listadas na Figura 18.

Variável	Descrição
Sexo	Feminino ou masculino
Idade	Idade do cliente no dia do cadastro (em anos)
Estado Civil	Casado, solteiro, divorciado, viúvo, etc.
Escolaridade	Fundamental, médio, superior incompleto ou completo
Renda	Valor da renda (R\$)
Tipo de Renda	Renda declarada ou comprovada
Profissão	Profissão ou cargo do cliente
Tipo Ocupação	Assalariado, autônomo, profissional liberal, etc.
CEP Residencial	CEP do local onde reside
CEP Comercial	CEP do local onde trabalha
Tempo Serviço	Tempo no emprego atual (em meses)
Crédito 3ºs	Tem crédito em outros estabelecimentos?
Tipo Residência	Própria, alugada, cedida ou com pais
Cidade Nascimento	Cidade de naturalidade do cliente
Filho	Tem filhos?
Pensão	Paga pensão alimentícia?

Figura 18 Variáveis identificadas para criação do modelo

Fonte: proposta de crédito da empresa (2004)

O período de cadastro dos clientes constante na amostra é limitado, já que é necessária a espera de 12 a 18 meses após a inclusão do cliente para verificar a definição de seu desempenho como pagador. Desta forma, a sazonalidade existente na quantidade de cadastros incluídos não foi totalmente contemplada. O período do inverno, que apresenta um aumento na busca pelo crédito em decorrência das doenças da estação, não fez parte da amostra em sua totalidade. O período amostrado compreende os cadastros de clientes aprovados de dezembro de 2005 a junho de 2006. O total de clientes na amostra e a quantidade por tipo de cliente são apresentados na Tabela 3.

Tabela 3
Total de clientes por tipo

Tipo Cliente	Quantidade	%
Mau	3655	21,5
Bom	8026	47,2
Indefinido	795	4,7
Sem Uso	4529	26,6
Total	17005	100

Como somente os clientes classificados como bons e maus são considerados para o desenvolvimento do modelo, a amostra considerada neste estudo teve um total de 11.681 clientes. A utilização de uma amostra de tamanho grande, nesse caso, garante uma reprodução mais confiável dos inter-relacionamentos estabelecidos entre as variáveis do modelo.

Antes de passar para a separação da amostra total em duas partes (análise e teste) é importante uma avaliação geral do preenchimento dos dados para eliminar dados inconsistentes, *missing* e *outliers*. Analisando as informações de idade e tempo de serviço, verificou-se a ocorrência de valores negativos e também casos de clientes com menos de 18 anos, demonstrando erros de preenchimento nos campos. Na informação de CEP residencial, observaram-se casos de CEP geral (por exemplo: 90.000-000 ou 91.000-000), zerados ou de outros estados. Todos estes casos foram eliminados, fazendo que com que a amostra de trabalho ficasse com um total de 11.394 clientes.

De forma aleatória, foram separadas as amostras de análise e teste na proporção de 80% e 20%, respectivamente. A amostra de teste, reservada para a posterior validação dos modelos construídos, selecionada aleatoriamente, ficou composta de 1631 clientes bons e 648 maus, num total de 2279 clientes.

A amostra de análise ficou formada, portanto, por 6305 clientes bons e 2720 maus. Desta forma, foi utilizado para a construção dos modelos um total de 9115 observações. Esta quantidade está além do proposto por Lewis (1992) de 1500 casos para cada grupo de cliente bom e mau. Segundo Hair *et al.* (2005), a amostra também é suficiente, tendo um número total maior que o sugerido que é de pelo menos 20 observações para cada variável independente (20 x 69 variáveis *dummies* = 1380 casos). A obtenção das 69 variáveis *dummies* é explicada na próxima seção.

4.3 ANÁLISE PRELIMINAR

A análise para escolha das variáveis é realizada através do cálculo de risco relativo, dividindo-se o percentual de bons clientes pelo percentual de maus de cada atributo. Nesta fase inicial, duas variáveis foram excluídas da análise por terem poder de discriminação muito baixo, evidenciado pelo valor do risco relativo bem próximo de 1, como pode ser observado na Tabela 4.

Tabela 4
Risco relativo de variáveis excluídas da análise

VARIÁVEL	Mau	Bom	RR
Tipo Renda			
Comprovada	1472	3717	1,07
Declarada	1248	2678	0,91
Crédito 3ºs			
Sim	1068	2608	1,04
Não	1652	3787	0,98
Total	2720	6395	9115

A variável com a informação de pagamento de pensão alimentícia também foi excluída da análise por ter poucos casos de clientes com esta característica (2,13% - 194 casos). A renda é outra variável que não foi levada em consideração na construção do modelo pelo fato de não haver um comportamento linear quanto ao valor do risco relativo (Tabela 5).

Tabela 5
Risco relativo da variável salário

Salário	Mau	Bom	Total	RR	Classe Risco
até 300	42	100	142	1,01	Neutro
300 -- 400	382	1159	1541	1,29	Bom
400 -- 500	415	902	1317	0,92	Neutro
500 -- 600	317	740	1057	0,99	Neutro
600 -- 700	372	734	1106	0,84	Mau
700 -- 800	177	398	575	0,96	Neutro
800 -- 900	361	804	1165	0,95	Neutro
900 -- 1000	155	271	426	0,74	Mau
1000 -- 1200	169	377	546	0,95	Neutro
1200 -- 1400	107	263	370	1,05	Neutro
1400 -- 1600	96	205	301	0,91	Neutro
1600 -- 2000	54	170	224	1,34	Bom
2000 -- 3000	48	166	214	1,47	Bom
3000 ou +	25	106	131	1,80	Muito Bom
Total	2720	6395	9115		

Para que as variáveis profissão, cidade de nascimento, CEP residencial e comercial fizessem parte da análise foi necessário agrupá-las, dado o grande número de atributos de cada uma delas. Para tal agrupamento foi utilizado o resultado do risco relativo de cada atributo juntando os que tivessem resultados próximos, segundo a escala: péssimo (RR < 0,50); muito mau (RR entre 0,50 e 0,67); mau (RR entre 0,67 e 0,90); neutro (RR entre 0,90 e 1,10); bom (RR entre 1,10 e 1,50); muito bom (RR entre 1,50 e 2,00) e excelente (RR maior que 2,00). Para que fosse possível proceder a classificação, foi estipulada a ocorrência mínima de 30 observações em cada atributo. Um exemplo do procedimento de classificação é apresentado na Tabela 6.

Tabela 6
Exemplo de agrupamento de atributos segundo risco relativo

Profissão	Mau	Bom	Total	RR	Classificação
Agricultor	11	40	51	1,55	Muito Bom
Ajudante	6	8	14	0,57	Não Classificado
Aposentado	262	1397	1659	2,27	Excelente
Artesão	10	9	19	0,38	Não Classificado
Atendente	22	55	77	1,06	Neutro
Aux. cozinha	12	24	36	0,85	Mau
Aux. limpeza	3	13	16	1,84	Não Classificado
Aux. produção	30	40	70	0,57	Muito Mau
Aux. serviços gerais	130	228	358	0,75	Mau

Para proceder com o agrupamento das variáveis CEP residencial e comercial foi preciso desagregar a informação por passos. Como a informação de CEP é composta de 8 dígitos, é necessário analisar os atributos partindo dos dois primeiros dígitos e seguindo para a análise de três e quatro dígitos, sempre que for possível ter uma representação mínima de 30 casos. Por exemplo, se o total de casos com o CEP residencial inicial 910 for representativo, analisa-se a partição de CEP de 9100 a 9109.

Desta forma, o agrupamento das variáveis com muitos atributos, segundo a análise do risco relativo, levou a criação de sete grupos, do cliente péssimo ao excelente. Cada grupo de risco é transformado em uma variável *dummy* (0 ou 1) que serão, portanto, as variáveis independentes para a construção do modelo. O resultado para as quatro variáveis em questão é apresentado na Tabela 7.

Tabela 7
Agrupamento final e variáveis *dummies* segundo risco relativo

VARIÁVEL	Dummy	Mau	Bom	Total	RR
Grupos de Profissões					
Péssimo	DGPROF1	90	83	173	0,39
Muito Mau	DGPROF2	364	502	866	0,59
Mau	DGPROF3	883	1623	2506	0,78
Neutro	DGPROF4	254	566	820	0,95
Bom	DGPROF5	184	505	689	1,17
Muito Bom	DGPROF6	92	375	467	1,73
Excelente	DGPROF7	282	1528	1810	2,30
Grupos de Cidades de Nascimento					
Péssimo	DCIDNA1	73	79	152	0,46
Muito Mau	DCIDNA2	1075	1478	2553	0,58
Mau	DCIDNA3	387	733	1120	0,81
Neutro	DCIDNA4	253	569	822	0,96
Bom	DCIDNA5	181	543	724	1,28
Muito Bom	DCIDNA6	127	507	634	1,70
Excelente	DCIDNA7	72	585	657	3,46
Grupos de CEP Residencial					
Péssimo	DGCEPRE1	18	17	35	0,40
Muito Mau	DGCEPRE2	559	824	1383	0,63
Mau	DGCEPRE3	916	1709	2625	0,79
Neutro	DGCEPRE4	237	544	781	0,98
Bom	DGCEPRE5	513	1536	2049	1,27
Muito Bom	DGCEPRE6	80	339	419	1,80
Excelente	DGCEPRE7	88	718	806	3,47
Grupos de CEP Comercial					
Péssimo	DGCEPCO1	40	38	78	0,40
Muito Mau	DGCEPCO2	328	472	800	0,61
Mau	DGCEPCO3	520	958	1478	0,78
Neutro	DGCEPCO4	56	140	196	1,06
Bom	DGCEPCO5	218	639	857	1,25
Muito Bom	DGCEPCO6	24	89	113	1,58
Excelente	DGCEPCO7	102	695	797	2,90

As relações dos atributos classificados em cada um dos sete grupos para as variáveis profissão, cidade de nascimento, CEP residencial e comercial são apresentados no Apêndice A, Apêndice B, Apêndice C e Apêndice D, respectivamente.

Para as demais informações do banco de dados também foram criadas variáveis *dummies*. As categorias de escolaridade – superior incompleto e superior completo – foram agrupadas em uma só categoria (superior) para que ficassem todas em um mesmo nível de desagregação. As variáveis idade e tempo de serviço foram segmentadas em classes, também segundo análise do risco relativo. Os resultados são apresentados na Tabela 8.

Tabela 8
Criação de variáveis *dummies* para demais variáveis independentes

VARIÁVEL	Dummy	Mau	Bom	Total	RR	Classe de Risco
Idade						
ATÉ 20 anos	DIDAD1	457	406	863	0,38	Péssimo
21 a 25 anos	DIDAD2	421	656	1077	0,66	Muito Mau
26 a 30 anos	DIDAD3	438	629	1067	0,61	Muito Mau
31 a 35 anos	DIDAD4	344	604	948	0,75	Mau
36 a 40 anos	DIDAD5	283	656	939	0,99	Neutro
41 a 50 anos	DIDAD6	411	1261	1672	1,30	Bom
51 a 60 anos	DIDAD7	206	1035	1241	2,14	Excelente
acima de 60 anos	DIDAD8	143	1108	1251	3,30	Excelente
Sexo						
Feminino	DSEXOF	1748	4550	6298	1,11	Bom
Masculino	DSEXOM	972	1845	2817	0,81	Mau
Escolaridade						
Fundamental	DPRI	1344	3206	4550	1,01	Neutro
Médio	DSEC	1217	2631	3848	0,92	Neutro
Superior	DSUP	159	558	717	1,49	Bom
Estado Civil						
Casado	DCASADO	585	2193	2778	1,59	Muito Bom
Concubinado	DCONCUB	20	45	65	0,96	Neutro
Divorciado	DDIVOR	148	445	593	1,28	Bom
Outros	DOUTR	161	375	536	0,99	Neutro
Separado	DSEPARA	74	163	237	0,94	Neutro
Solteiro	DSOLTE	1593	2500	4093	0,67	Muito Mau
Viúvo	DVIUVO	139	674	813	2,06	Excelente
Tempo de Serviço						
até 3 meses	DTSERV1	43	94	137	0,93	Neutro
4 a 6 meses	DTSERV2	94	123	217	0,56	Muito Mau
7 a 18 meses	DTSERV3	262	412	674	0,67	Muito Mau
19 a 24 meses	DTSERV4	75	131	206	0,74	Mau
25 a 36 meses	DTSERV5	105	192	297	0,78	Mau
37 a 60 meses	DTSERV6	79	227	306	1,22	Bom
61 a 90 meses	DTSERV7	39	158	197	1,72	Muito Bom
91 a 120 meses	DTSERV8	13	93	106	3,04	Excelente
121 meses ou +	DTSERV9	45	264	309	2,50	Excelente
Tem Filho?						
Não	DNFILHO	1683	4357	6040	1,10	Bom
Sim	DFILHO	1037	2038	3075	0,84	Mau
Tipo de Ocupação						
Aposentado	DOCUP_AP	342	1645	1987	2,05	Excelente
Assalariado	DOCUP_AS	1153	2350	3503	0,87	Mau
Autônomo	DOCUP_AU	1122	2009	3131	0,76	Mau
Funcionário público	DOCUP_FP	60	290	350	2,06	Excelente
Profissional liberal	DOCUP_PL	43	101	144	1,00	Neutro
Tipo de Residência						
Reside com pais	DRES_PAI	53	62	115	0,50	Péssimo
Alugada	DRES_ALU	325	531	856	0,69	Mau
Cedida	DRES_CED	181	339	520	0,80	Mau
Outras	DRES_OUT	308	565	873	0,78	Mau
Própria	DRES_PRO	1853	4898	6751	1,12	Bom

4.4 CONSTRUÇÃO DO MODELO PRC

Os resultados da construção dos modelos de previsão de risco de crédito com utilização das técnicas de análise discriminante, regressão logística e redes neurais são apresentados na seqüência.

Para a construção dos modelos com uso das técnicas de análise discriminante e regressão logística, foi utilizado o SPSS versão 13.0 (*Statistical Package for Social Science*). O *software* utilizado para treinamento e teste das redes neurais, empregada para a construção do terceiro modelo proposto, foi o BrainMaker Professional versão 3.7.

4.4.1 Análise Discriminante

Nos testes iniciais para construção do modelo, através do método *stepwise*, utilizaram-se parâmetros para a entrada e saída de variáveis do modelo de 5% e 10%, respectivamente. Dentre as 69 variáveis *dummies*, relacionadas como possíveis variáveis independentes, apenas 33 tiveram poder discriminatório significativo para compor o modelo final, como é apresentado na Tabela 9.

Tabela 9
Coefficientes da função discriminante canônica

Variáveis	Coefficientes	Significância
DIDAD1	-1,172	0,000
DIDAD23	-0,547	0,000
DIDAD4	-0,315	0,008
DIDAD6	0,272	0,009
DIDAD7	0,650	0,000
DIDAD8	0,933	0,000
DSEXOF	0,366	0,000
DPRIM	-0,344	0,000
DSUP	0,257	0,016
DCASADO	0,436	0,000
DTSERV67	0,492	0,000
DTSERV89	0,641	0,000
DFILHO	-0,343	0,000
DRES_ALU	-0,580	0,000
DGCEPR12	-0,602	0,000
DGCEPRE3	-0,329	0,000
DGCEPCO1	-0,949	0,001
DGCEPC56	0,298	0,001
DGCEPCO7	0,707	0,000
DGPROF1	-0,936	0,000
DGPROF2	-0,399	0,000
DGPROF5	0,212	0,049
DGPROF67	0,293	0,000
DCIDNA12	-0,613	0,000
DCIDNA3	-0,406	0,000
DCIDNA7	0,457	0,000
(Constant)	0,154	0,000

Para que algumas variáveis tivessem significância para entrar no modelo final, foi necessário fazer o agrupamento de *dummies* próximas, como por exemplo, os grupos de cidade de nascimento 1 e 2 (péssimo e muito mau) e os grupos de

profissões 6 e 7 (muito bom e excelente), entre outros. Com tais ajustes, o modelo discriminante final foi obtido com 26 variáveis independentes.

A interpretação da equação discriminante demonstra a ponderação atribuída a cada atributo para a separação dos clientes nos grupos. O sinal dos coeficientes de cada uma das variáveis indica o sentido para a classificação do tipo de cliente, sendo um indicativo de uma característica para um cliente mau o sinal negativo, e de um cliente bom o sinal positivo. Ao multiplicar os valores de cada variável independente para cada cliente pelos respectivos coeficientes, encontra-se o escore discriminante.

Obtido o modelo através da análise discriminante, parte-se para sua avaliação e validação. Na Tabela 10 é apresentada a distribuição dos clientes bons, maus e total de acordo com o escore do modelo e também as taxas de sinistro.

Tabela 10
Distribuição de clientes e taxa de sinistro do modelo discriminante

CLASSE ESCORE DISCRIMINANTE	MAUS		BONS		TOTAL		TAXA DE SINISTRO	
	#	%	#	%	#	%	CLASSE	TOTAL
-4,0 a -3,5	1	0,0%	1	0,0%	2	0,0%	50,0%	29,8%
-3,5 a -3,0	6	0,2%	3	0,0%	9	0,1%	66,7%	29,8%
-3,0 a -2,5	55	2,0%	17	0,3%	72	0,8%	76,4%	29,8%
-2,5 a -2,0	126	4,6%	55	0,9%	181	2,0%	69,6%	29,4%
-2,0 a -1,5	374	13,8%	207	3,2%	581	6,4%	64,4%	28,6%
-1,5 a -1,0	455	16,7%	485	7,6%	940	10,3%	48,4%	26,1%
-1,0 a -0,5	527	19,4%	680	10,6%	1207	13,2%	43,7%	23,2%
-0,5 a 0,0	479	17,6%	992	15,5%	1471	16,1%	32,6%	19,2%
0,0 a 0,5	354	13,0%	1160	18,1%	1514	16,6%	23,4%	15,0%
0,5 a 1,0	185	6,8%	1142	17,9%	1327	14,6%	13,9%	10,9%
1,0 a 1,5	119	4,4%	1012	15,8%	1131	12,4%	10,5%	8,7%
1,5 a 2,0	32	1,2%	441	6,9%	473	5,2%	6,8%	5,7%
2,0 a 2,5	6	0,2%	150	2,3%	156	1,7%	3,8%	3,4%
2,5 a 3,0	1	0,0%	40	0,6%	41	0,4%	2,4%	2,0%
3,0 a 3,5	0	0,0%	7	0,1%	7	0,1%	0,0%	0,0%
3,5 a 4,0	0	0,0%	3	0,0%	3	0,0%	0,0%	0,0%
TOTAL	2720	100%	6395	100%	9115	100%	-	-

A coluna da taxa de sinistro total informa o percentual de maus clientes que seriam aprovados sobre o total de clientes aprovados se fosse utilizado como corte o limite inferior da classe. Por exemplo, se fosse utilizado como corte o escore -2, esperar-se-ia 28,6% de maus clientes no total aprovado. A taxa de sinistro na classe informa ainda que na classe de escore de -2 a -1,5 tem 64,4% de clientes maus, quase 2/3 dos clientes nesta faixa. O que se observa é um decréscimo nesta taxa

que é explicado pelo poder de separação entre os bons e maus clientes, que pode também ser observado na Figura 19.

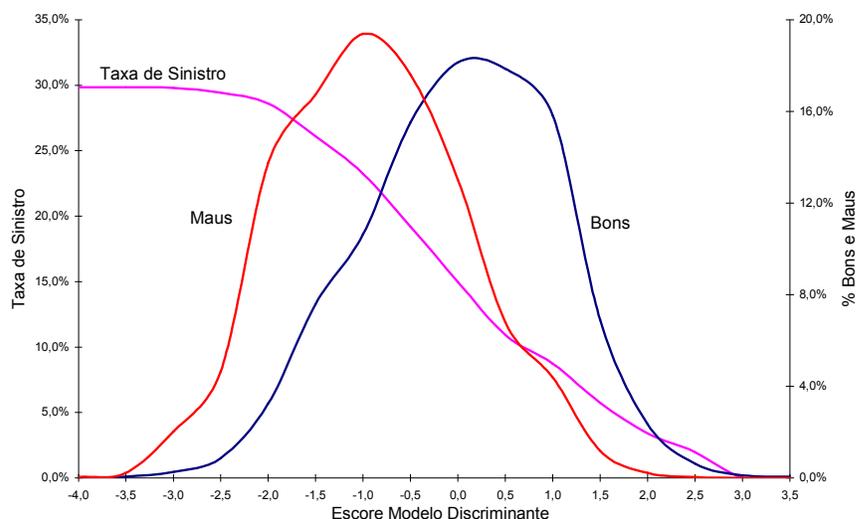


Figura 19 Taxa de sinistro e separação do modelo discriminante

A distribuição dos escores do modelo discriminante para os grupos de bons e maus clientes permite identificar a tendência de que os clientes maus têm escores negativos e que os bons têm escores positivos. O resultado desta separação é visualizado com a queda na taxa de sinistro conforme se avança na direção dos maiores escores. É importante salientar que a separação obtida entre as duas curvas é considerada pequena e que tal resultado é devido ao número e tipo de variáveis utilizadas na análise. Incorporar outras variáveis em uma avaliação futura do modelo deve melhorar o afastamento dos perfis.

Uma questão importante a se verificar diz respeito ao comportamento dos escores do modelo nas duas amostras, de análise e teste, ou seja, um primeiro passo de verificação de adequação do modelo encontrado. Espera-se que as distribuições dos escores nas duas amostras não sejam diferentes, e para testar utiliza-se o mesmo teste KS para duas amostras que será útil para verificar o poder de previsão do modelo e a comparação com os demais modelos construídos. Os resultados desta comparação são apresentados na Tabela 11.

Tabela 11
Comparação entre amostras segundo modelo discriminante

CLASSE DE ESCORE DISCRIMINANTE	AMOSTRA DE ANÁLISE				AMOSTRA DE TESTE				TESTE KS %
	TOTAL		ACUMULADO		TOTAL		ACUMULADO		
	#	%	#	%	#	%	#	%	
-4,0 a -3,5	2	0,02%	2	0,02%	1	0,04%	1	0,04%	0,02%
-3,5 a -3,0	9	0,10%	11	0,12%	2	0,09%	3	0,13%	0,01%
-3,0 a -2,5	72	0,79%	83	0,91%	11	0,48%	14	0,61%	-0,30%
-2,5 a -2,0	181	1,99%	264	2,90%	47	2,06%	61	2,68%	-0,22%
-2,0 a -1,5	581	6,38%	845	9,27%	141	6,19%	202	8,86%	-0,41%
-1,5 a -1,0	940	10,32%	1785	19,59%	224	9,83%	426	18,69%	-0,90%
-1,0 a -0,5	1207	13,25%	2992	32,84%	307	13,47%	733	32,16%	-0,67%
-0,5 a 0,0	1471	16,14%	4463	48,98%	372	16,32%	1105	48,49%	-0,49%
0,0 a 0,5	1514	16,62%	5977	65,59%	371	16,28%	1476	64,77%	-0,83%
0,5 a 1,0	1327	14,56%	7304	80,16%	347	15,23%	1823	79,99%	-0,17%
1,0 a 1,5	1131	12,41%	8435	92,57%	286	12,55%	2109	92,54%	-0,03%
1,5 a 2,0	473	5,19%	8908	97,76%	125	5,48%	2234	98,03%	0,26%
2,0 a 2,5	156	1,71%	9064	99,47%	35	1,54%	2269	99,56%	0,09%
2,5 a 3,0	41	0,45%	9105	99,92%	9	0,39%	2278	99,96%	0,03%
3,0 a 3,5	7	0,08%	9112	100,00%	1	0,04%	2279	100,00%	0,00%
3,5 a 4,0	3	0,03%	9115	100,03%	0	0,00%	2279	100,00%	-0,03%
TOTAL	9112	100%	9112	100%	2279	100%	2279	100%	0,90%
Valor Crítico =>									3,19%

A avaliação dos percentuais de clientes revela uma distribuição dos escores muito semelhante nas duas amostras, já que as diferenças são muito pequenas, como esperado. O reflexo disso está no baixo valor de KS encontrado (0,90) que é bem menor que o valor crítico do teste (3,19), o que leva a aceitar a hipótese de igualdade de distribuições.

Para o emprego do modelo encontrado na previsão do risco de crédito com clientes da empresa é necessário verificar os pressupostos da análise discriminante. Com a utilização do método *stepwise* garantiu-se que o pressuposto da inexistência de multicolinearidade fosse atendido, já que o método prioriza a inclusão de variáveis independentes com alto poder discriminatório e também que sejam menos correlacionadas entre si.

A suposição de normalidade é verificada com a aplicação do teste de Kolmogorov-Smirnov (KS) para uma amostra. Este teste compara as frequências observadas com as esperadas segundo a distribuição normal. Quanto maior o valor do teste, maior é a distância entre a distribuição dos dados e a normal. O valor obtido no teste foi de 0,031 e significativo ($p=0,000$) o que leva a decisão de rejeitar a hipótese de ajuste à distribuição normal. Porém esta significância pode ser explicada

pela alta sensibilidade do teste, em função do grande tamanho da amostra. Ainda assim, o ajuste aproximadamente normal pode ser observado no formato do histograma apresentado na Figura 20.

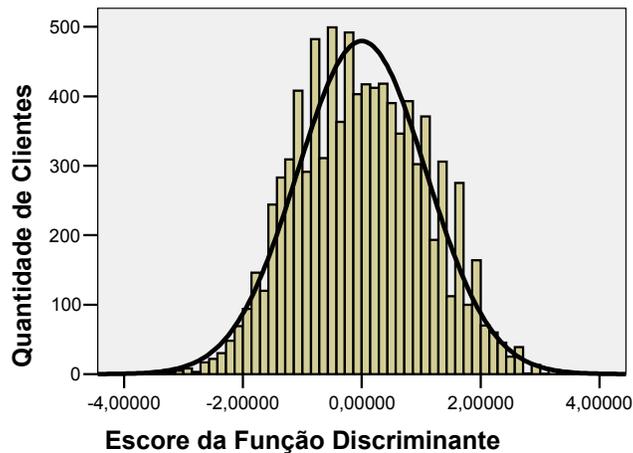


Figura 20 Distribuição dos escores da função discriminante

O último pressuposto a ser verificado é o da homogeneidade de matrizes de variâncias e covariâncias dos grupos bom e mau. Tal premissa, necessária para a utilização da técnica, é verificada através do teste denominado Box's M. O resultado do teste para o modelo construído é apresentado na Tabela 12.

Tabela 12
Resultado do teste Box's M

Box's M	8275,970
F aproximado	23,492
df1	351
df2	91342202
Significância	0,000

O resultado do teste indica que as matrizes de variâncias e covariâncias dos dois grupos não são semelhantes, uma vez que a hipótese de igualdade é rejeitada (significância = 0,000). Porém, Hair *et al.* (2005) salientam que o teste Box' M é extremamente sensível ao tamanho da amostra. Por essa razão, a análise do modelo obtido para fins de previsão não fica prejudicada, já que o objetivo aqui é comparar o poder de predição dos modelos encontrados.

4.4.2 Regressão Logística

Assim como na construção do modelo pela análise discriminante, com a regressão logística também se utilizou o método *stepwise* com parâmetros de significância para a entrada de variáveis do modelo de 5% e de saída de 10%. Dentre as 69 variáveis *dummies* apresentadas como possíveis variáveis independentes, 35 foram significativas para compor o modelo final, como é apresentado na Tabela 13.

Tabela 13
Coefficientes da regressão logística

Variáveis	B	Sig.	Exp(B)
DIDAD1	-0,829	0,000	0,436
DIDAD23	-0,409	0,000	0,665
DIDAD4	-0,252	0,014	0,778
DIDAD6	0,232	0,014	1,261
DIDAD7	0,644	0,000	1,904
DIDAD8	1,047	0,000	2,849
DSEXOF	0,327	0,000	1,386
DPRIM	-0,287	0,000	0,750
DSUP	0,270	0,009	1,310
DCASADO	0,410	0,000	1,507
DTSERV6	0,340	0,017	1,405
DTSERV7	0,627	0,001	1,871
DTSERV89	0,792	0,000	2,208
DFILHO	-0,293	0,000	0,746
DRES_ALU	-0,547	0,000	0,579
DGCEPR12	-0,392	0,000	0,676
DGCEPRE3	-0,172	0,015	0,842
DGCEPRE5	0,197	0,016	1,217
DGCEPRE6	0,328	0,028	1,388
DGCEPRE7	0,608	0,000	1,837
DGCEPCO1	-0,768	0,002	0,464
DGCEPC56	0,218	0,016	1,243
DGCEPCO7	0,472	0,003	1,603
DGPROF1	-0,718	0,000	0,488
DGPROF2	-0,318	0,000	0,727
DGPROF67	0,283	0,000	1,327
DCIDNA12	-0,449	0,000	0,638
DCIDNA3	-0,328	0,000	0,720
DCIDNA7	0,592	0,000	1,808
Constant	0,876	0,000	2,401

Algumas variáveis tiveram que ser agrupadas, assim como na análise discriminante, para que se verificasse a significância estatística e elas fizessem parte do modelo final. A coluna Exp(B), que é o coeficiente de cada variável elevado na constante e (2,7182...), pode ser interpretada como o impacto de cada coeficiente sobre a razão de chance. Por exemplo, ser casado aumenta em 50,7% ($e^{0,41}=1,507$) a chance de ser bom cliente, supondo que as demais variáveis se mantenham constantes.

Com o agrupamento de algumas variáveis *dummies* para atingir significância, o modelo final construído incluiu 29 variáveis independentes. De acordo com os coeficientes encontrados, o modelo logístico resultante é apresentado na Equação 06.

$$P(Y=1) = \frac{1}{1 + \exp(0,876 - 0,829 \text{ DIDAD1} - 0,409 \text{ DIDAD23} - 0,252 \text{ DIDAD4} + 0,232 \text{ DIDAD6} + 0,644 \text{ DIDAD7} + 1,047 \text{ DIDAD8} + 0,327 \text{ DSEXOF} - 0,287 \text{ DPRIM} + 0,270 \text{ DSUP} + 0,410 \text{ DCASADO} + 0,340 \text{ DTSERV6} + 0,627 \text{ DTSERV7} + 0,792 \text{ DTSERV89} - 0,293 \text{ DFILHO} - 0,547 \text{ DRES_ALU} - 0,392 \text{ DGCEPR12} - 0,172 \text{ DGCEPRE3} + 0,197 \text{ DGCEPRE5} + 0,328 \text{ DGCEPRE6} + 0,608 \text{ DGCEPRE7} - 0,768 \text{ DGCEPCO1} + 0,218 \text{ DGCEPC56} + 0,472 \text{ DGCEPCO7} - 0,718 \text{ DGPROF1} - 0,318 \text{ DGPROF2} + 0,283 \text{ DGPROF67} - 0,449 \text{ DCIDNA12} - 0,328 \text{ DCIDNA3} + 0,592 \text{ DCIDNA7})} \quad (06)$$

A probabilidade de um proponente vir a ser um bom cliente é dada pela Equação 06. Desta forma, as variáveis com sinais positivos revelam associações com ser bom pagador e as de sinais negativos com ser mau pagador. Ou seja, um proponente que tem idade acima de 41 anos, tem curso superior, é casado, tem mais de 3 anos de tempo de serviço, entre outras, tem maior probabilidade de ser um bom pagador.

Partindo para a avaliação do poder de separação do modelo construído, primeiramente, verifica-se a distribuição dos clientes de acordo com a probabilidade obtida pelo modelo e também as taxas de sinistro por classe e total, apresentado na Tabela 14.

Tabela 14
Distribuição de clientes e taxa de sinistro do modelo logístico

CLASSE PROB. REG. LOGÍSTICA	MAUS		BONS		TOTAL		TAXA DE SINISTRO	
	#	%	#	%	#	%	CLASSE	TOTAL
0,1 - 0,15	2	0,1%	2	0,0%	4	0,0%	50,0%	29,8%
0,15 - 0,2	8	0,3%	6	0,1%	14	0,2%	57,1%	29,8%
0,2 - 0,25	32	1,2%	8	0,1%	40	0,4%	80,0%	29,8%
0,25 - 0,3	73	2,7%	25	0,4%	98	1,1%	74,5%	29,6%
0,3 - 0,35	127	4,7%	60	0,9%	187	2,1%	67,9%	29,1%
0,35 - 0,4	193	7,1%	103	1,6%	296	3,2%	65,2%	28,2%
0,4 - 0,45	211	7,8%	167	2,6%	378	4,1%	55,8%	27,0%
0,45 - 0,5	229	8,4%	220	3,4%	449	4,9%	51,0%	25,6%
0,5 - 0,55	246	9,0%	304	4,8%	550	6,0%	44,7%	24,1%
0,55 - 0,6	291	10,7%	363	5,7%	654	7,2%	44,5%	22,5%
0,6 - 0,65	251	9,2%	397	6,2%	648	7,1%	38,7%	20,3%
0,65 - 0,7	243	8,9%	488	7,6%	731	8,0%	33,2%	18,2%
0,7 - 0,75	236	8,7%	606	9,5%	842	9,2%	28,0%	16,1%
0,75 - 0,8	195	7,2%	653	10,2%	848	9,3%	23,0%	13,7%
0,8 - 0,85	156	5,7%	798	12,5%	954	10,5%	16,4%	11,3%
0,85 - 0,9	137	5,0%	1050	16,4%	1187	13,0%	11,5%	9,4%
0,9 - 0,95	76	2,8%	866	13,5%	942	10,3%	8,1%	7,3%
0,95 - 1,0	14	0,5%	279	4,4%	293	3,2%	4,8%	4,8%
TOTAL	2720	100%	6395	100%	9115	100%	-	-

Na coluna da taxa de sinistro total é informado o percentual de maus clientes que seriam aprovados se fosse utilizado como corte o valor inferior da classe de probabilidade em questão. A taxa de sinistro na classe é o percentual de clientes maus nesta faixa. Pode-se observar um decréscimo nas taxas de sinistro que é reflexo da separação entre os bons e maus clientes obtida pelo modelo, que pode também ser observado na Figura 21.

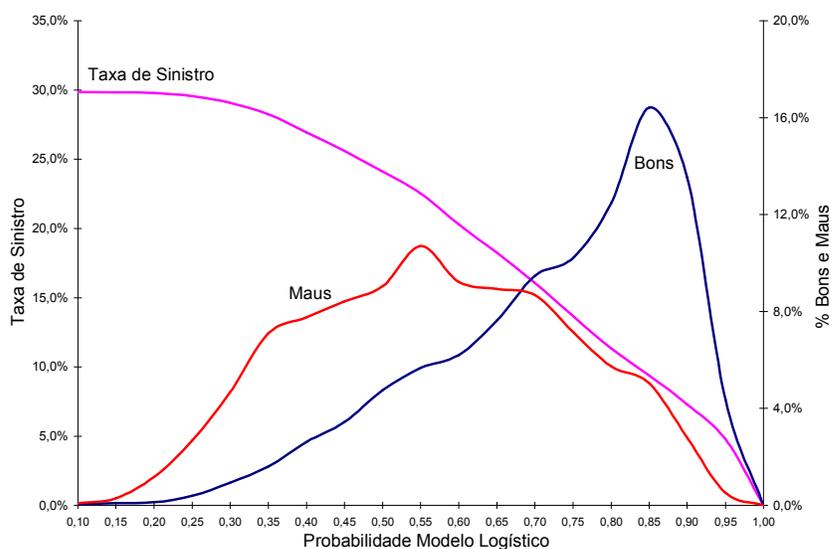


Figura 21 Taxa de sinistro e separação do modelo logístico

Considerando a distribuição das probabilidades do modelo logístico, é possível verificar a tendência de que os maus clientes estão mais para a esquerda da escala (probabilidade menor que 50%) e que os bons para a direita (probabilidade maior que 50%). Esta separação é observada também com a queda da taxa de sinistro, partindo de 30% e chegando a quase 0%, nos maiores valores de probabilidade.

Um passo importante na verificação de adequação do modelo logístico encontrado é a avaliação do comportamento das probabilidades nas duas amostras, de análise e teste. O teste KS para verificar a homogeneidade da distribuição dos resultados do modelo nas amostras é apresentado na Tabela 15.

Tabela 15
Comparação entre amostras segundo modelo logístico

CLASSE DE PROBABILIDADE M. LOGÍSTICO	AMOSTRA DE ANÁLISE				AMOSTRA DE TESTE				TESTE KS %
	TOTAL		ACUMULADO		TOTAL		ACUMULADO		
	#	%	#	%	#	%	#	%	
0,1 - 0,15	4	0,04%	4	0,04%	1	0,04%	1	0,04%	0,00%
0,15 - 0,2	14	0,15%	18	0,20%	3	0,13%	4	0,18%	-0,02%
0,2 - 0,25	40	0,44%	58	0,64%	5	0,22%	9	0,39%	-0,24%
0,25 - 0,3	98	1,08%	156	1,71%	22	0,97%	31	1,36%	-0,35%
0,3 - 0,35	187	2,05%	343	3,76%	43	1,89%	74	3,25%	-0,52%
0,35 - 0,4	296	3,25%	639	7,01%	64	2,81%	138	6,06%	-0,96%
0,4 - 0,45	378	4,15%	1017	11,16%	96	4,21%	234	10,27%	-0,89%
0,45 - 0,5	449	4,93%	1466	16,08%	111	4,87%	345	15,14%	-0,95%
0,5 - 0,55	550	6,03%	2016	22,12%	137	6,01%	482	21,15%	-0,97%
0,55 - 0,6	654	7,17%	2670	29,29%	164	7,20%	646	28,35%	-0,95%
0,6 - 0,65	648	7,11%	3318	36,40%	167	7,33%	813	35,67%	-0,73%
0,65 - 0,7	731	8,02%	4049	44,42%	183	8,03%	996	43,70%	-0,72%
0,7 - 0,75	842	9,24%	4891	53,66%	208	9,13%	1204	52,83%	-0,83%
0,75 - 0,8	848	9,30%	5739	62,96%	206	9,04%	1410	61,87%	-1,09%
0,8 - 0,85	954	10,47%	6693	73,43%	247	10,84%	1657	72,71%	-0,72%
0,85 - 0,9	1187	13,02%	7880	86,45%	304	13,34%	1961	86,05%	-0,40%
0,9 - 0,95	942	10,33%	8822	96,79%	261	11,45%	2222	97,50%	0,71%
0,95 - 1,0	293	3,21%	9115	100,00%	57	2,50%	2279	100,00%	0,00%
TOTAL	9115	100%	9115	100%	2279	100%	2279	100%	1,09%

Valor Crítico => **3,19%**

A distribuição dos percentuais de clientes nas classes de probabilidades nas duas amostras é muito semelhante. Esta homogeneidade de distribuição é verificada pelo baixo valor do teste KS (1,09) que é menor que o valor crítico do teste (3,19).

No que diz respeito à verificação de atendimento de premissas, somente a ausência de multicolinearidade deve ser confirmada. Com a utilização do método *stepwise* para escolha das variáveis independentes para compor o modelo, garantiu-se o atendimento deste pressuposto.

4.4.3 Redes Neurais

O *software* utilizado para a construção das redes neurais não tem nenhum tipo de método como o *stepwise*, utilizado na obtenção dos modelos anteriores, que seleciona automaticamente as variáveis com maior poder de explicação. Portanto, utilizou-se para o estudo das redes as variáveis que foram indicadas previamente através dos modelos discriminante e logístico, e também por um teste exploratório feito com a regressão linear.

Para a obtenção das redes neurais utilizou-se a função de ativação sigmóide e o algoritmo de aprendizado supervisionado de retropropagação de erro, com somente uma camada escondida. Várias redes foram criadas com diferentes quantidades de neurônios na camada escondida para verificar o desempenho quanto à predição dos bons e maus clientes. Para avaliar o desempenho das redes compararam-se os valores do erro quadrático médio e do teste KS para as amostras de análise e de teste. Os resultados das melhores redes construídas são apresentados na Tabela 16.

Tabela 16
Comparação dos melhores modelos neurais construídos

Modelo	Nº Neurônios camada oculta	KS	
		Análise	Teste
Log13	35	38,3	34,0
Dis13	30	39,2	33,9
Reg15	30	39,3	35,1
Dis12	35	41,0	32,7
Reg13	27	40,3	34,6
Dis11	35	40,3	35,4

O modelo neural escolhido, com melhor desempenho para previsão do risco de crédito foi o modelo Dis11. O modelo Dis12 teve melhores resultados para a amostra de análise, porém o desempenho na amostra de teste é bem menor, evidenciando o excesso de encaixe da rede. Os pesos obtidos para os neurônios da camada oculta e da camada de saída são apresentados no Apêndice E.

A aplicação do modelo neural não é tão simples quanto o emprego dos modelos discriminante e logístico, dado suas relações não-lineares entre as variáveis. Porém, é possível replicar o modelo de rede neural através de uma planilha eletrônica e utilizá-lo para previsões. No Anexo A é apresentada a formulação para réplica das redes utilizada por Oliveira (2004) em seu trabalho com redes neurais.

Após a escolha do modelo neural com melhor desempenho e da implementação em planilha eletrônica, passa-se para sua avaliação e validação. Na Tabela 17 é apresentada a distribuição dos clientes bons, maus e total de acordo com o score do modelo e também as taxas de sinistro.

Tabela 17
Distribuição de clientes e taxa de sinistro do modelo neural

CLASSE ESCORE NEURAL	MAUS		BONS		TOTAL		TAXA DE SINISTRO	
	#	%	#	%	#	%	CLASSE	TOTAL
0,05 - 0,1	2	0,1%	1	0,0%	3	0,0%	66,7%	29,8%
0,1 - 0,15	5	0,2%	0	0,0%	5	0,1%	100,0%	29,8%
0,15 - 0,2	10	0,4%	4	0,1%	14	0,2%	71,4%	29,8%
0,2 - 0,25	24	0,9%	5	0,1%	29	0,3%	82,8%	29,7%
0,25 - 0,3	78	2,9%	13	0,2%	91	1,0%	85,7%	29,6%
0,3 - 0,35	85	3,1%	24	0,4%	109	1,2%	78,0%	29,0%
0,35 - 0,4	116	4,3%	42	0,7%	158	1,7%	73,4%	28,4%
0,4 - 0,45	112	4,1%	48	0,8%	160	1,8%	70,0%	27,6%
0,45 - 0,5	164	6,0%	97	1,5%	261	2,9%	62,8%	26,8%
0,5 - 0,55	158	5,8%	114	1,8%	272	3,0%	58,1%	25,6%
0,55 - 0,6	182	6,7%	167	2,6%	349	3,8%	52,1%	24,5%
0,6 - 0,65	239	8,8%	235	3,7%	474	5,2%	50,4%	23,3%
0,65 - 0,7	214	7,9%	294	4,6%	508	5,6%	42,1%	21,5%
0,7 - 0,75	293	10,8%	480	7,5%	773	8,5%	37,9%	19,9%
0,75 - 0,8	269	9,9%	495	7,7%	764	8,4%	35,2%	17,6%
0,8 - 0,85	224	8,2%	659	10,3%	883	9,7%	25,4%	14,9%
0,85 - 0,9	166	6,1%	766	12,0%	932	10,2%	17,8%	12,8%
0,9 - 0,95	209	7,7%	1189	18,6%	1398	15,3%	14,9%	11,4%
0,95 - 1,0	170	6,3%	1762	27,6%	1932	21,2%	8,8%	8,8%
TOTAL	2720	100%	6395	100%	9115	100%	-	-

A taxa de sinistro na classe informa o percentual de clientes maus nesta faixa, como por exemplo, na faixa de escore de 0,65 a 0,7 tem-se que 42,1% dos clientes são maus pagadores. Já a taxa de sinistro total é o percentual de maus clientes que seriam aprovados se fosse utilizado como corte o valor inferior da classe de probabilidade em questão. Por exemplo, se o ponto de corte adotado for de 0,5, espera-se que 25,6% dos clientes aprovados venham a se tornar maus clientes. O decréscimo nas taxas de sinistro que é reflexo da separação entre os bons e maus clientes obtida pelo modelo pode também ser observado na Figura 22.

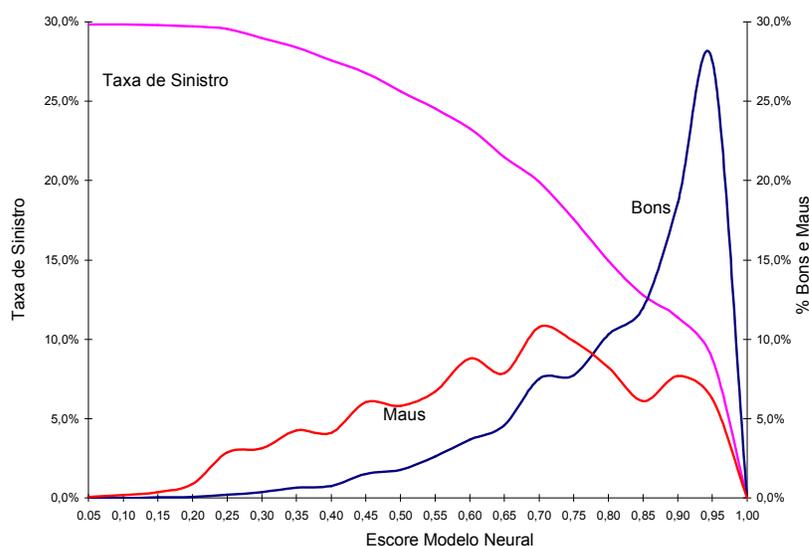


Figura 22 Taxa de sinistro e separação do modelo neural

Analisando o comportamento das curvas de distribuição dos bons e maus clientes, verifica-se que o modelo neural consegue separar os dois grupos de clientes, já que é possível observar a tendência de que os maus clientes estão mais para a esquerda da escala e que os bons para a direita. A queda na taxa de sinistro, à medida que aumenta o valor dos escores, também é um reflexo da separação obtida pelo modelo.

Uma última validação necessária ao modelo neural encontrado trata-se da verificação do comportamento dos escores nas duas amostras, de análise e de teste. Através desta verificação espera-se comprovar uma homogeneidade da distribuição dos resultados dos escores nas duas amostras, necessária para utilização do modelo como previsão do risco de crédito a novos clientes. Na Tabela 18 é apresentado o resultado do teste KS que é utilizado para testar esta homogeneidade das distribuições.

Tabela 18
Comparação entre amostras segundo modelo neural

CLASSE DE ESCORE M. NEURAL	AMOSTRA DE ANÁLISE				AMOSTRA DE TESTE				TESTE KS %
	TOTAL		ACUMULADO		TOTAL		ACUMULADO		
	#	%	#	%	#	%	#	%	
0,05 - 0,1	3	0,03%	3	0,03%	0	0,00%	0	0,00%	-0,03%
0,1 - 0,15	5	0,05%	8	0,09%	1	0,04%	1	0,04%	-0,04%
0,15 - 0,2	14	0,15%	22	0,24%	2	0,09%	3	0,13%	-0,11%
0,2 - 0,25	29	0,32%	51	0,56%	6	0,26%	9	0,39%	-0,16%
0,25 - 0,3	91	1,00%	142	1,56%	9	0,39%	18	0,79%	-0,77%
0,3 - 0,35	109	1,20%	251	2,75%	23	1,01%	41	1,80%	-0,95%
0,35 - 0,4	158	1,73%	409	4,49%	40	1,76%	81	3,55%	-0,93%
0,4 - 0,45	160	1,76%	569	6,24%	47	2,06%	128	5,62%	-0,63%
0,45 - 0,5	261	2,86%	830	9,11%	61	2,68%	189	8,29%	-0,81%
0,5 - 0,55	272	2,98%	1102	12,09%	70	3,07%	259	11,36%	-0,73%
0,55 - 0,6	349	3,83%	1451	15,92%	86	3,77%	345	15,14%	-0,78%
0,6 - 0,65	474	5,20%	1925	21,12%	110	4,83%	455	19,96%	-1,15%
0,65 - 0,7	508	5,57%	2433	26,69%	137	6,01%	592	25,98%	-0,72%
0,7 - 0,75	773	8,48%	3206	35,17%	202	8,86%	794	34,84%	-0,33%
0,75 - 0,8	764	8,38%	3970	43,55%	169	7,42%	963	42,26%	-1,30%
0,8 - 0,85	883	9,69%	4853	53,24%	229	10,05%	1192	52,30%	-0,94%
0,85 - 0,9	932	10,22%	5785	63,47%	254	11,15%	1446	63,45%	-0,02%
0,9 - 0,95	1398	15,34%	7183	78,80%	336	14,74%	1782	78,19%	-0,61%
0,95 - 1,0	1932	21,20%	9115	100,00%	497	21,81%	2279	100,00%	0,00%
TOTAL	9115	100%	9115	100%	2279	100%	2279	100%	1,30%

Valor Crítico => **3,19%**

Pode-se verificar, portanto, que a distribuição dos percentuais de clientes nas classes de probabilidades nas duas amostras é muito semelhante. Esta

homogeneidade de distribuição é verificada pelo baixo valor do teste KS (1,30), apresentado na última coluna, que é menor que o valor crítico do teste (3,19).

4.5 ESCOLHA DO MODELO

Um meio adequado de verificar o poder de previsão dos modelos construídos é a medição do percentual de classificações corretas. Para definir a classificação dos clientes é necessário definir um ponto de corte na escala de escores, acima do qual é classificado como bom cliente e abaixo do qual como mau.

Como os tamanhos dos dois grupos são diferentes, na análise discriminante faz-se necessário obter um ponto de corte ponderado, que neste caso foi de -1,12. Utilizando tal valor como regra para classificação dos clientes, obteve-se os percentuais de acertos de classificação para avaliação da qualidade do modelo discriminante que são apresentados na Tabela 19.

Tabela 19
Percentuais de acerto do modelo discriminante

AMOSTRA	GRUPOS ORIGINAIS	CLASSIFICAÇÃO		
		MAU	BOM	TOTAL
ANÁLISE	MAU	872 (32,1%)	1848 (67,9%)	2720
	BOM	598 (9,4%)	5797 (90,6%)	6395
	TOTAL	1470 (16,1%)	7645 (83,9%)	9115
TESTE	MAU	179 (27,6%)	469 (72,4%)	648
	BOM	167 (10,2%)	1464 (89,8%)	1631
	TOTAL	346 (15,2%)	1933 (84,8%)	2279

A porcentagem de acertos é uma informação importante, pois reflete a qualidade de ajuste do modelo de previsão de risco de crédito. A Tabela 19 revela que 872 casos do grupo dos maus clientes e 5797 casos do grupo dos bons clientes foram classificados corretamente, perfazendo o total de 6669 casos corretamente classificados, o que corresponde a 73,2% do total, com uma aprovação de 83,9% dos proponentes. O valor encontrado é superior ao percentual que se teria ao classificar todos os casos no grupo maior, os bons, que representam 70,2%. A classificação de todos os clientes no grupo dos bons é equivalente à aprovação total dos proponentes, ou seja, sem discriminação alguma dos grupos.

Na amostra de teste o percentual total de classificações corretas foi de 72,1%, com 27,6% dos maus e 89,8% dos bons pagadores bem classificados. Desta forma, a aprovação total obtida nesta amostra foi de 84,8%.

Utilizando-se um ponto de corte de 0,5 tem-se, na Tabela 20, a matriz das classificações dos clientes a posteriori e sua comparação com as classificações a priori segundo o modelo logístico para as amostras de análise e teste.

Tabela 20
Percentuais de acerto do modelo logístico

AMOSTRA	GRUPOS ORIGINAIS	CLASSIFICAÇÃO		
		MAU	BOM	TOTAL
ANÁLISE	MAU	875 (32,2%)	1845 (67,8%)	2720
	BOM	591 (9,2%)	5804 (90,8%)	6395
	TOTAL	1466 (16,1%)	7649 (83,9%)	9115
TESTE	MAU	180 (27,8%)	468 (72,2%)	648
	BOM	165 (10,1%)	1466 (89,9%)	1631
	TOTAL	345 (15,1%)	1934 (84,9%)	2279

O percentual de acerto total do modelo logístico foi de 73,3% para a amostra de análise, com 875 maus e 5804 bons corretamente classificados. Na amostra de teste esse percentual foi de 72,2%, sendo 180 maus e 1466 bons corretamente classificados. Os percentuais de aprovação para as duas amostras também foram bem próximos, com 83,9% para a amostra de análise e de 84,9% para a amostra de teste.

Assim como na análise discriminante, no modelo neural obtém-se também um ponto de corte, sendo que nesta aplicação o valor 0,6 mostrou-se mais adequado (a literatura não apresenta sugestão de ponte de corte na aplicação de redes neurais). Na Tabela 21 é apresentada a matriz das classificações dos clientes com os percentuais de acertos segundo o modelo neural para as amostras de análise e teste.

Tabela 21
Percentuais de acerto do modelo neural

AMOSTRA	GRUPOS ORIGINAIS	CLASSIFICAÇÃO		
		MAU	BOM	TOTAL
ANÁLISE	MAU	936 (34,4%)	1784 (65,6%)	2720
	BOM	515 (8,1%)	5880 (91,9%)	6395
	TOTAL	1451 (15,9%)	7664 (84,1%)	9115
TESTE	MAU	240 (37,0%)	408 (63,0%)	648
	BOM	215 (13,2%)	1416 (86,8%)	1631
	TOTAL	455 (20,0%)	1824 (80,0%)	2279

Para o modelo neural, o percentual de acerto total na classificação para a amostra de análise foi de 74,8%, com 936 maus e 5880 bons corretamente classificados, chegando a uma aprovação de 84,1% do total de clientes. Na amostra de teste o percentual de classificações corretas foi de 72,7%, sendo 240 maus e 1416 bons corretamente classificados, atingindo uma aprovação total de 80%.

Para avaliar os desempenhos reais das taxas de aprovações dos modelos é necessário verificar os percentuais de classificação dos modelos em todos os clientes da base de dados, incluindo os grupos indefinido e sem uso. Na Tabela 22 são apresentados os percentuais de classificações como mau e bom cliente segundo os três modelos para os quatro grupos de clientes. Desta forma, pode-se verificar que os percentuais de aprovações dos modelos ficam em 83,9%, 84,1%, 83,5% para os modelos discriminante, logístico e neural, respectivamente, adotando os pontos de corte previamente estabelecidos.

Tabela 22
Classificação dos grupos segundo os modelos construídos

GRUPOS ORIGINAIS	DISCRIMINANTE		LOGÍSTICO		NEURAL		TOTAL
	Negar (%)	Aprovar (%)	Negar (%)	Aprovar (%)	Negar (%)	Aprovar (%)	
MAU	1122 (30,7)	2533 (69,3)	1121 (30,7)	2534 (69,3)	1188 (32,5)	2467 (67,5)	3655
BOM	776 (9,7)	7250 (90,3)	761 (9,5)	7265 (90,5)	768 (9,6)	7258 (90,4)	8026
INDEFINIDO	112 (14,1)	683 (85,9)	112 (14,1)	683 (85,9)	111 (14,0)	684 (86,0)	795
SEM USO	727 (16,1)	3802 (83,9)	717 (15,8)	3812 (84,2)	737 (16,3)	3793 (83,7)	4529
TOTAL	2737 (16,1)	14268 (83,9)	2711 (15,9)	14294 (84,1)	2804 (16,5)	14201 (83,5)	17005

Outra forma de avaliar a qualidade de um modelo de previsão de risco de crédito é através do teste não-paramétrico Kolmogorov-Smirnov (KS). Este teste tem por objetivo determinar se duas amostras provêm de uma mesma população. No presente caso, espera-se provar que as duas amostras de clientes (bons e maus) provêm de populações distintas o que significaria que o modelo consegue separar os dois grupos. A comparação dos valores do teste KS para os três modelos construídos é apresentada na Tabela 23.

Tabela 23
Valores de KS para os modelos construídos

Modelo	Amostra	
	Análise	Teste
Discriminante	36,7	30,7
Logístico	36,9	31,7
Neural	40,3	35,4

Todos os valores encontrados para KS são maiores que 30, valor mínimo para considerar um modelo com bom poder de separação. Observa-se novamente que o resultado obtido pela regressão logística foi praticamente idêntico ao da análise discriminante. A separação obtida com o modelo neural foi ligeiramente superior aos demais modelos, que pode ser explicada por sua abordagem diferenciada no relacionamento das variáveis.

Na Figura 23 são representados os valores de KS para os três modelos construídos para as amostras de análise. Através dos gráficos é possível visualizar o grau de separação entre os clientes maus e bons através dos escores dos modelos.

De acordo com as duas medidas de desempenho usadas aqui, percentual total de classificações corretas e do valor do teste de KS, conclui-se que o modelo que teve maior poder de discriminação entre os grupos de clientes bons e maus foi o neural, sendo, portanto, o indicado para implantação na empresa.

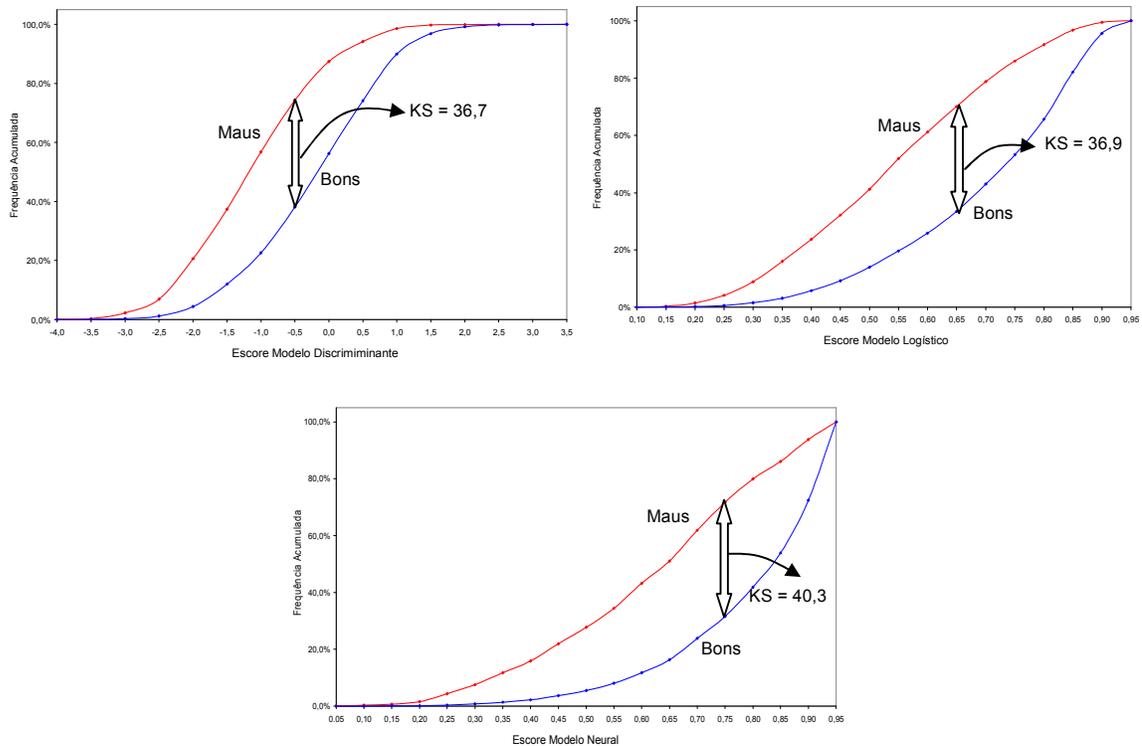


Figura 23 Representação do valor de KS para os modelos construídos

4.6 PASSOS PARA IMPLANTAÇÃO

Para programar o modelo de previsão de risco de crédito obtido no sistema da empresa são necessárias algumas diretrizes sobre sua parametrização. É imprescindível que os campos de profissão e cidade de nascimento sejam fechados com códigos para escolha dos atributos, evitando, com isso, problemas de mau preenchimento que acarretariam na não pontuação da característica pelo modelo.

Também é importante, para o bom funcionamento do modelo, que as variáveis que o compõem tenham preenchimento obrigatório, garantindo sua devida ponderação para análise do risco do cliente. Na medida do possível, o mesmo poderia ser feito com outros campos que não fizeram parte do modelo neste momento, podendo, em uma nova reavaliação, serem considerados.

A determinação do ponto de corte é um passo mais gerencial que metodológico, já que envolve questões importantes para a empresa como taxa de aprovação e proporção de clientes bons negados. Para nortear essa decisão, uma análise de sensibilidade de ponto de corte para os modelos discriminante, logístico e neural é apresentada na Tabela 24, Tabela 25 e Tabela 26, respectivamente.

Tabela 24
Sensibilidade do ponto de corte do modelo discriminante

PONTO DE CORTE	% de Acerto			% de Aprovação
	Mau	Bom	Total	
-4	0,0%	100,0%	70,2%	100,0%
-3,5	0,0%	100,0%	70,2%	100,0%
-3	0,3%	99,9%	70,2%	99,9%
-2,5	2,3%	99,7%	70,6%	99,1%
-2	6,9%	98,8%	71,4%	97,1%
-1,5	20,7%	95,6%	73,2%	90,7%
-1	37,4%	88,0%	72,9%	80,4%
-0,5	56,8%	77,4%	71,2%	67,2%
0	74,4%	61,8%	65,6%	51,0%
0,5	87,4%	43,7%	56,7%	34,4%
1	94,2%	25,8%	46,2%	19,9%
1,5	98,6%	10,0%	36,4%	7,5%
2	99,7%	3,1%	32,0%	2,3%
2,5	100,0%	0,8%	30,4%	0,6%
3	100,0%	0,2%	30,0%	0,1%
3,5	100,0%	0,0%	29,9%	0,0%

Verifica-se que o ponto de corte que maximiza o percentual de acerto total está em torno de -1,5 pontos, tendo-se com isso uma taxa de aprovação de 90,7% já que 20,7% dos maus pagadores estariam sendo negados e também 4,4% dos bons. Para um estudo mais aprofundado do ponto de corte ideal seria interessante entender o comportamento dos percentuais de acertos entre os valores de escore -1,5 e -1,0 já que é visível que é neste intervalo que o percentual de acerto total é maximizado.

Para que seja mais fácil a compreensão e a implantação do modelo, sugere-se, para o modelo discriminante, que todos os coeficientes do modelo sejam multiplicados por 100 e arredondados para inteiros para facilitar a interpretação dos coeficientes pela empresa. No caso da constante, também é interessante que seja aumentada para 500 para que não haja a possibilidade de se ter escores negativos, fato que na prática pode vir a causar problemas.

Tabela 25
Sensibilidade do ponto de corte do modelo logístico

PONTO DE CORTE	% de Acerto			% de Aprovação
	Mau	Bom	Total	
0,1	0,0%	100,0%	70,2%	100,0%
0,15	0,1%	100,0%	70,2%	100,0%
0,2	0,4%	99,9%	70,2%	99,8%
0,25	1,5%	99,7%	70,4%	99,4%
0,3	4,2%	99,4%	71,0%	98,3%
0,35	8,9%	98,4%	71,7%	96,2%
0,4	16,0%	96,8%	72,7%	93,0%
0,45	23,8%	94,2%	73,2%	88,8%
0,5	32,2%	90,8%	73,3%	83,9%
0,55	41,2%	86,0%	72,6%	77,9%
0,6	51,9%	80,3%	71,8%	70,7%
0,65	61,1%	74,1%	70,2%	63,6%
0,7	70,1%	66,5%	67,6%	55,6%
0,75	78,8%	57,0%	63,5%	46,3%
0,8	85,9%	46,8%	58,5%	37,0%
0,85	91,7%	34,3%	51,4%	26,6%
0,9	96,7%	17,9%	41,4%	13,5%
0,95	99,5%	4,4%	32,7%	3,2%
1	100,0%	0,0%	29,8%	0,0%

Para o modelo logístico, o ponto de corte que maximiza o percentual de acerto total na classificação é 0,5 com 32,2% dos maus pagadores sendo reprovados e 9,2% dos bons, chegando a uma taxa de aprovação de 83,9%.

Tabela 26
Sensibilidade do ponto de corte do modelo neural

PONTO DE CORTE	% de Acerto			% de Aprovação
	Mau	Bom	Total	
0,05	0,0%	100,0%	70,2%	100,0%
0,1	0,1%	100,0%	70,2%	100,0%
0,15	0,3%	100,0%	70,2%	99,9%
0,2	0,6%	99,9%	70,3%	99,8%
0,25	1,5%	99,8%	70,5%	99,4%
0,3	4,4%	99,6%	71,2%	98,4%
0,35	7,5%	99,3%	71,9%	97,2%
0,4	11,8%	98,6%	72,7%	95,5%
0,45	15,9%	97,9%	73,4%	93,8%
0,5	21,9%	96,3%	74,1%	90,9%
0,55	27,7%	94,6%	74,6%	87,9%
0,6	34,4%	91,9%	74,8%	84,1%
0,65	43,2%	88,3%	74,8%	78,9%
0,7	51,1%	83,7%	73,9%	73,3%
0,75	61,8%	76,2%	71,9%	64,8%
0,8	71,7%	68,4%	69,4%	56,4%
0,85	80,0%	58,1%	64,6%	46,8%
0,9	86,1%	46,1%	58,1%	36,5%
0,95	93,8%	27,6%	47,3%	21,2%
1	100,0%	0,0%	29,8%	0,0%

O ponto de corte no modelo neural que maximiza o percentual de classificações corretas fica em torno de 0,6 com 74,8%. Adotando-se este ponto de corte, tem-se 84,1% de aprovação com 34,4% dos maus sendo negados e 8,1% dos bons.

Um proponente a crédito deverá sofrer apenas uma análise pelo modelo, de maneira a garantir o sigilo e a eficiência da fórmula, ou seja, devem-se evitar simulações ou testes com o modelo. A proposta não deverá sofrer intervenção alguma quanto ao preenchimento dos campos. Nos casos em que alguma informação não foi declarada ou recuperada, deverão ser mantidos em branco.

Outra forma de utilizar os escores obtidos com o modelo de previsão de risco de crédito é o gerenciamento da concessão do crédito. Como as pontuações dão um ordenamento dos clientes de maior risco até os de menor risco, é possível utilizá-las para, por exemplo, decidir pela aprovação ou liberação de valores menores de limite, escolher clientes para participar de alguma promoção, flexibilizar a exigência por alguns documentos, entre outros.

Dado que no novo processo de decisão de crédito o modelo implantado será peça fundamental e que ele se baseia no pressuposto que as informações declaradas são verídicas, recomenda-se o acesso restrito às informações que compõem a fórmula e seus pesos, bem como o ponto de corte, ao menor número de pessoas possível, evitando-se fraudes no processo de crédito.

Uma forma de reduzir os gastos relativos a consultas aos órgãos de restrição de crédito é utilizar previamente o modelo de previsão de risco de crédito. Desta forma, as consultas aos clientes negados pelo modelo tornam-se desnecessárias, refletindo uma economia para a empresa. O fluxo da análise de crédito com utilização do Modelo PRC construído é apresentada na Figura 24.

A indicação de realização de consultas aos órgãos de restrição de crédito é devido ao alto percentual de clientes maus pagadores que são aprovados pelos modelos, chegando em torno de 65% de aprovação, como pode ser observado na Tabela 22. Esse alto percentual de classificações incorretas dos clientes maus como bons pagadores é a justificativa para o caráter conservador do fluxo proposto.

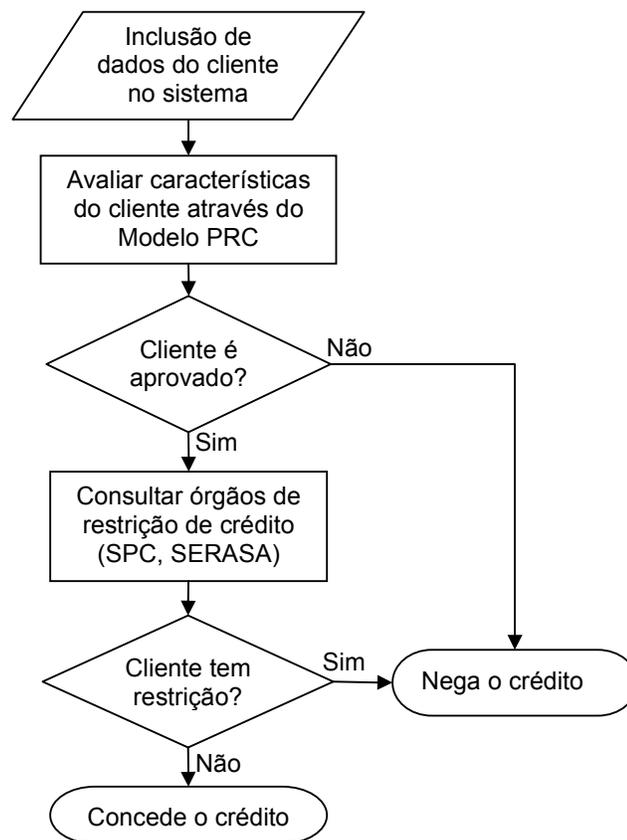


Figura 24 Fluxograma para análise de crédito usando o Modelo PRC

Para que o modelo seja empregado como ferramenta de decisão, são necessários relatórios periódicos para acompanhamento de sua adequada utilização, verificando o perfil demográfico dos clientes que estão solicitando crédito. Este perfil será confrontado com o perfil da amostra de desenvolvimento do modelo e não deve ter alterações significativas, o que pode ser sinal de alterações de informações para interferir na pontuação e aprovação dos clientes. Para testar a homogeneidade dos perfis, pode-se utilizar o teste Qui-quadrado de homogeneidade para comparação de duas distribuições. O cálculo do teste Qui-quadrado é feito segundo a Equação 07 e deve ser menor que o valor crítico com $n-1$ graus de liberdade e 5% de significância.

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i} \quad (07)$$

onde,

n = número de atributos da variável;

o = percentual observado para o atributo no mês de escoragem;

e = percentual esperado para o atributo segundo a amostra de desenvolvimento.

No Apêndice F, é apresentado o formato do relatório para confrontar o perfil dos clientes da amostra de desenvolvimento (denominado perfil padrão) com o perfil atual dos clientes cadastrados, mês a mês. Valores altos para o Qui-quadrado em alguma característica podem ser reflexo de dois tipos de condições que devem ser investigadas e tratadas adequadamente.

Diferenças de distribuição dos perfis podem ser evidências de alterações nas características das propostas dos proponentes com o objetivo de beneficiar sua pontuação e atingir a aprovação como cliente. Neste caso, a empresa deve analisar as propostas de crédito e validar com as informações cadastradas no sistema ou até mesmo contatar os clientes para confirmação.

Outra possibilidade de alteração nos perfis dos clientes pode ser devido a algum tipo de captação direcionada (por exemplo: promoção especial a aposentados) ou mudanças demográficas graduais (por exemplo: migração, envelhecimento, aumento de salário). É importante que, para estes casos, sejam monitorados os comportamentos de pagamento, não devendo haver trocas de perfis, como por exemplo, aposentados tornarem-se maus pagadores. Se houver alterações deste tipo, sugere-se a reavaliação do sistema através da construção de um novo modelo.

4.7 AVALIAÇÃO GERENCIAL

A etapa de avaliação gerencial do Modelo PRC teve como objetivo verificar a aplicabilidade dos modelos criados na empresa em estudo. Os resultados alcançados foram apresentados a um grupo de pessoas envolvidas com o

gerenciamento do processo de concessão do crédito e da área de tecnologia da informação da empresa. O Modelo PRC foi apresentado em sua totalidade e em detalhes, desde as definições iniciais para coleta da amostra até os passos para a implementação no sistema. Desta forma, a empresa tem o conhecimento necessário para novas reavaliações do modelo.

Apresentados os três modelos construídos, a empresa demonstrou interesse em implementar o modelo neural em seu sistema, devido ao seu melhor desempenho em relação aos outros modelos. Apesar de uma aparente complexidade de implementação do modelo, a área de TI considerou bastante útil e didático o material de réplica da rede neural exposto no Anexo A. Segundo a empresa, o tempo estimado de implementação será de quatro meses, contemplando as seguintes etapas:

- inclusão de procedimentos de validação de campos necessários e correção de banco de cadastros antigos;
- determinação de pontos de corte, fluxo de aprovação e liberação de crédito (incluindo ponderações de limites);
- desenvolvimento da solução no sistema;
- testes e pareceres internos;
- implantação de pilotos em algumas lojas;
- treinamento;
- disponibilização final para rede de lojas;
- utilização dos escores para reavaliação de limites, focando clientes potencialmente bons.

A empresa apontou como principais vantagens: (i) seleção de clientes para a concessão do cartão; (ii) redução de custos relativos à consulta ao SPC; (iii) ponderação para liberação de limites de crédito. Acreditam ainda que o modelo possa ser utilizado como forma de comparação da qualidade da análise de crédito

feita em cada loja. Com isso, será possível estabelecer um indicador de meta para as lojas como forma de direcionar as premiações e treinamentos.

Para a utilização do modelo construído, foi sugerido que o fluxo de concessão de crédito, apresentado na Figura 24, seja utilizado também para redução de gastos de consultas aos órgãos restritivos dos proponentes com maiores escores. Desta forma, não só se deixaria de consultar no SPC os clientes abaixo do ponto de corte (negados pelo modelo), mas também os clientes com baixa possibilidade de inadimplência (com escore maior que um ponto estipulado pela empresa).

4.8 AVALIAÇÃO DE GANHO POTENCIAL

A avaliação do ganho potencial foi realizada somente para o modelo com melhor desempenho, o modelo neural. Para tal avaliação, primeiramente estudou os valores médios de compras feitas no período analisado, levando em consideração os clientes considerados na amostra. Tais valores são apresentados na Tabela 27.

Tabela 27
Média de compras dos clientes no período

Tipo de Cliente	Nº Clientes	Nº de compras	Valor de compras
Mau	3655	9,1	252,43
Bom	8026	17,1	463,45
Indefinido	795	14,9	395,54
Sem Uso	4529	0,0	0,00
Total	17005	14,7	291,49

Utilizando-se o modelo neural com o ponto de corte sugerido, espera-se uma aprovação total de 83,5% dos clientes, sendo que 67,5% dos maus, 86% dos indefinidos e 90,4% dos bons serão aprovados, como foi apresentado na Tabela 22. Considerando-se, por exemplo, que o recebimento da venda realizada a clientes bons é de 100%, aos indefinidos de 80% e aos maus de 40% e, ainda, que a margem líquida da empresa é de 15%, obtêm-se os resultados de venda e ganho apresentados na Tabela 28.

Tabela 28
Estimativas de venda e ganho para clientes da amostra

	Sem Modelo	Com Modelo
Venda Total	4.956.750,48	4.255.786,74
Investimento 85% da venda total	4.213.237,91	3.617.418,73
Recebimento bons (100%), indefinidos (80%) e maus (40%)	4.340.288,21	3.828.039,85
Lucro recebimento - investimento	127.050,31	210.621,12

Observa-se que, com a utilização do modelo neural, espera-se uma queda de 14% na venda total. Porém o percentual de recebimento da venda total que sem o modelo é de 87,5% aumenta para 89,9% com a utilização do modelo neural, o que se reflete em aumento no lucro esperado que é dado pela diferença entre o recebimento e o investimento (percentual da venda total referente aos custos de mercadoria, operacional e impostos).

Considerando, portanto, a receita relativa à base de clientes estudada, esperar-se-ia, com o uso do modelo neural, um aumento de lucratividade de R\$ 83.570,81 (R\$210.621,12 – R\$127.050,31), ou seja, um acréscimo de 65,8% nos lucros referente às vendas aos clientes da amostra em estudo.

4.9 DISCUSSÃO DOS RESULTADOS

Com a aplicação da sistemática proposta, alguns resultados alcançados podem ser discutidos à luz do caso específico em estudo. Segundo Lewis (1992), a definição do desempenho de pagamento considerado como satisfatório pela empresa pode influenciar o poder de separação do modelo construído. Em alguns setores, um atraso de 30 dias pode ser considerado aceitável (natural); em outros, pode caracterizar o mau pagador. Portanto, atenção especial neste sentido é recomendada na etapa inicial de construção do Modelo PRC. O critério que separa bons e maus pagadores deve ser definido tendo em vista o setor específico (negócio) que está sendo considerado.

Os resultados obtidos com a aplicação da técnica de rede neural demonstraram que o poder de separação entre os grupos de clientes bons e maus pagadores, conforme critérios definidos pela empresa, foi levemente superior aos resultados encontrados com as demais técnicas abordadas. Outros autores também se propuseram a mesma comparação entre as técnicas multivariadas na construção de modelos de previsão de risco de crédito, obtendo poder de previsão semelhante entre os modelos construídos com as diferentes técnicas (DESAI *et al.*, 1996; ARMINGER *et al.*, 1997; MARQUES; LIMA, 2002).

Outra questão importante para o aumento da qualidade de previsão do Modelo PRC diz respeito ao número e a qualidade das variáveis independentes considerada. Um bom modelo de previsão de risco de crédito é reflexo direto da qualidade dos dados. Lawrence (1992) comenta sobre características utilizadas em modelos de previsão de risco de crédito e sobre a necessidade de inclusão dessas características na proposta da empresa para uma avaliação futura do modelo. Informações como seguro de carro, número de dependentes da renda, tempo na residência atual e informações de número de consultas em órgão de restrição de crédito (SPC e SERASA, no Brasil) são algumas das características citadas pelo autor.

Securato (1996) comenta sobre o dilema da decisão: com uma aprovação está-se assumindo algum nível de risco; com um não se elimina o risco, mas pode-se estar perdendo uma oportunidade. O modelo de previsão de risco de crédito deve ajudar o analista neste sentido, atribuindo uma probabilidade associada ao risco da concessão em questão. Porém, o erro de decisão sempre irá existir, o modelo pode indicar a aprovação de um cliente que não honrará com o débito e, também, poderá sugerir a negativa de outro que seria um bom pagador.

Para tentar reduzir esta aparente injustiça é importante associar a análise subjetiva à informação obtida com o modelo de previsão de risco de crédito para aumentar os acertos na decisão, diminuindo o número de clientes que estariam sendo incorretamente classificados. Desta forma, o analista, através da análise de informações adicionais, tentaria identificar os clientes bons que o modelo indicou como maus pagadores e sugeriu que tivessem o crédito negado e, principalmente,

reduzir o grande percentual de clientes maus pagadores que estariam sendo aprovados previamente.

Segundo Vasconcellos (2002) é importante que a utilização do modelo de previsão de risco de crédito não caracterize qualquer tipo de discriminação dos proponentes. Negar o crédito, alegando questões como raça ou sexo, é caracterizado como crime em determinados países. Por essa razão é importante a indicação de que as informações contidas no modelo sejam mantidas em sigilo. Qualquer conhecimento quanto às características que são consideradas para a obtenção do score final de um cliente pode levar o atendente a usá-las para justificar a negativa do crédito, o que poderia ser caracterizado como discriminatório.

O mesmo autor ainda ressalta a avaliação das flutuações amostrais que podem ser um problema para a aplicação do modelo de previsão de risco de crédito utilizado pela empresa. Com o passar do tempo alterações no perfil da população, como aumento de renda e modificações nos percentuais de pessoas casadas, por exemplo, podem levar a perda de desempenho da previsão do modelo. Por esse motivo, é sugerido o acompanhamento das características consideradas no modelo, sendo que alterações significativas indicam a necessidade de reavaliação do modelo, repetindo todas as etapas consideradas na sistemática proposta.

A avaliação da deterioração do modelo também pode ser realizada através do valor do Teste KS. Picinini *et al.* (2003) sugerem que a separação entre os grupos de bons e maus pagadores deve ser maior que 30%. Desta forma, avaliações periódicas, feitas após decorridos pelos 12 meses da concessão, podem evidenciar se o modelo construído necessita de uma nova reavaliação quando o valor de KS diminuir de 30.

5 CONSIDERAÇÕES FINAIS

Nesta seção são apresentadas as principais conclusões obtidas do estudo e algumas sugestões para trabalhos futuros.

5.1 CONCLUSÕES

O objetivo deste estudo foi construir um modelo de previsão de risco de crédito através da comparação das técnicas de análise discriminante, regressão logística e redes neurais. As características especiais de cada uma das técnicas foram analisadas com o objetivo de encontrar o modelo com melhor desempenho na predição de bons e maus clientes.

Todos os passos para a obtenção de um modelo de previsão de risco de crédito foram apresentados, desde a definição da população para aplicação até os passos para implementação no sistema da empresa. O nível de detalhe apresentado em cada etapa do método é inédito em trabalhos deste tipo. O Modelo PRC desenvolvido neste estudo pode, portanto, servir como instrumento de apoio para pesquisadores de empresas construírem seus modelos adaptados.

Os três modelos construídos tiveram desempenhos satisfatórios na previsão dos clientes como bons e maus pagadores, obtendo-se 73,2%, 73,3% e 74,8% de acerto na classificação com as técnicas de análise discriminante, regressão logística e redes neurais, respectivamente. Desta forma, observou-se uma superioridade da rede neural em relação às outras técnicas, explicada por sua abordagem não-linear na combinação das variáveis.

O modelo neural teve melhor separação na previsão que os outros dois modelos, verificado também pelo maior valor para o teste KS (40,3). Porém sua programação no sistema da empresa pode ser considerada complexa, o que poderia levar a empresa, numa decisão estratégica, a escolher o modelo discriminante ou o modelo logístico.

Os resultados obtidos com a análise discriminante e com a regressão logística foram muito semelhantes em termos de desempenho, também

evidenciados pelos valores do teste KS, 36,7 e 36,9, respectivamente. Desta forma, uma escolha entre os dois modelos seria dada através da análise dos pressupostos para utilização das técnicas, e neste caso, o modelo logístico deveria ser escolhido, por não ter nenhum dos pressupostos violados.

É importante lembrar que é necessário ter cuidado com casos de empresas estabelecidas em várias regiões, pois pode ser necessária a construção de modelos diferentes para cada região. No caso da empresa em estudo, suas atividades estão sediadas somente no estado do Rio Grande do Sul, o que limitou a construção de um único modelo.

A utilização dos modelos de previsão de risco de crédito elimina a subjetividade da análise, criando uma linha de decisão padrão que deve ser tomada como apoio, devendo a análise de crédito ser complementada com informações extras que não estejam contempladas no modelo. Desta forma é possível aumentar a velocidade da análise de crédito, o que pode permitir o aumento do número de clientes.

Os resultados obtidos foram apresentados à empresa que demonstrou interesse em implementar o modelo neural e adotá-lo como ferramenta de apoio em sua análise de crédito. Mesmo considerando complexa a implementação das redes neurais em comparação com os demais modelos, estima que em cinco meses tenha o sistema em funcionamento em todas as lojas da rede.

Com o intuito de demonstrar o ganho potencial do uso do modelo neural, estimou-se o aumento na lucratividade da empresa. Considerando a venda anual aos clientes da base em estudo, a empresa pode esperar um acréscimo de 65% nos lucros com o uso do modelo neural.

Como conclusão final, este estudo confirma que há diferentes técnicas que podem ser utilizadas para extrair informações relevantes tanto para a predição quanto ao pagamento de um crédito concedido. Cada técnica tem suas características e pressupostos que devem ser avaliados para que o modelo construído possa ser utilizado com confiança para fazer previsões do risco de crédito.

5.2 SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestão de trabalho futuro pode-se considerar a utilização do Modelo PRC, com algumas adaptações necessárias, como ferramenta de auxílio na atribuição de limites. Com o uso dos escores, que classificam os perfis de maior a menor risco, determinar percentuais da renda para liberação de limites de crédito

Outra sugestão é a utilização conjunta de técnicas estatísticas multivariadas, conhecida como *stacking*, permitindo uma melhora na predição através de um modelo único que incorpore várias técnicas para sua construção. Seria interessante também alterar o tratamento inicial das variáveis independentes, assumindo o valor do risco relativo ao invés do agrupamento deste na criação das variáveis *dummies*. Esta alteração pode refletir um ganho significativo no poder de predição do modelo, já que diminuiria a perda de informações existente quando do agrupamento em faixas de risco.

A mesma comparação de técnicas estatística multivariadas, feita neste estudo, poderia ser realizada para a construção de modelo comportamental (*Behavioural Scoring*) ou, ainda, verificar os resultados com a utilização de outras técnicas como Regressão Linear, Árvore de Decisão, Programação Linear, Análise de Sobrevivência, entre outras.

A comparação das técnicas também pode ser interessante em problemas de risco de crédito quando relacionada à concessão de crédito a pessoas jurídicas. Como se trata de um segmento diferenciado, é necessário um estudo mais aprofundado para verificar as características especiais que podem interferir na falência de uma empresa. A previsão de risco de crédito no pagamento da mensalidade de escolas e faculdades particulares também poderia ser outra aplicação do Modelo PRC que deve ter também suas características estudadas.

Como recomendação final, sugere-se a utilização e comparação das técnicas estatísticas multivariadas para criação de modelos de previsão em outras áreas do conhecimento. Poderia ser de interesse, por exemplo, saber se um empregado tende a se envolver em acidente no trabalho dada algumas características do seu ambiente de trabalho.

REFERÊNCIAS

ALMEIDA, F. C. Desvendando o uso de Redes Neurais em problemas de Administração de Empresas. *Revista de Administração de Empresas EAESP/FGV*, São Paulo, v.35, n.1, p.46-55, jan.-fev. 1995.

ALMEIDA, F. C.; DUMONTIER, P. O uso de Redes Neurais em avaliação de riscos de inadimplência. *Revista de Administração FEA/USP*, São Paulo, v.31, n.1, p.52-63, jan.-mar. 1996.

ALMEIDA, F. C.; SIQUEIRA, J. O. Comparação entre regressão logística e redes neurais na previsão de falência de bancos brasileiros. In: *Terceiro Congresso Brasileiro de Redes Neurais*, 1997, Florianópolis. Anais.

ALTMAN, E. I.; BAIDYA, T. K. N.; DIAS, L. M. R. Previsão de problemas financeiros em empresas. *Revista de Administração de Empresas*, São Paulo, v.19, n.1, p.19-28, jan.-mar. 1979.

ANDREEVA, G.; ANSELLA, J.; CROOK, J. Modelling profitability using survival combination scores. *European Journal of Operational Research*. v.183, n.3, p.1537-1549, dec. 2007.

ARAÚJO, E. A. Risco de crédito: Desenvolvimento de modelo Credit Scoring para a gestão da inadimplência de uma instituição de microcrédito. Monografia. ESAF – Escola de Administração Fazendária. IPEA – Instituto de Pesquisa Econômica Aplicada. 2006. Disponível em: <http://www.ipea.gov.br/ipeacaixa/premio2006/docs/trabpremiados/IpeaCaixa2006_Profissional_02lugar_tema03.pdf>. Acesso em 20 dez. 2007.

ARMINGER, G., ENACHE, D.; BONNE, T. Analyzing Credit Risk Data: A Comparison of Logistic Discrimination, Classification Trees and Feedforward Networks. *Computational Statistics*, v.12, n.2, p.293-310, 1997.

ASSAF NETO, A.; BRITO, G. Modelo de Classificação de Risco de Crédito de Grandes Empresas. In: *V Congresso USP de Controladoria e Contabilidade*, 2005, São Paulo. Anais.

BANCO CENTRAL DO BRASIL. *Fases de Crescimento do Crédito*. Relatório de Inflação, Set. 2007. Disponível em: < <http://www.bcb.gov.br/htms/relinf/port/2007/09/ri200709b2p.pdf>>. Acesso em: 05 jan. 2008.

BLATT, A. *Introdução a Análise de Crédito*. Rio de Janeiro: Suma Econômica, 1998.

BORGES, L. F. X.; BERGAMINI JUNIOR, S. O Risco Legal na Análise de Crédito. *Revista do BNES*, Rio de Janeiro, v.8, n.16, p.215-260, dez. 2001.

BUENO, V. F. F. *Avaliação de risco na concessão de crédito bancário para micros e pequenas empresas*. Florianópolis: UFSC, 2003. Dissertação (Mestrado em Engenharia da Produção), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 2003.

CAOQUETTE, J. B.; ALTMAN, E. I.; NARAYANAN, P. *Gestão do Risco de Crédito*. O próximo grande desafio financeiro. São Paulo: Qualitymark, 1999.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. *Análise Multivariada: para cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2007.

DESAI, V. S.; CROOK, J. N.; OVERSTREET JR., G. A. A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, v.95, n.1, p.24-37, nov. 1996.

EIFERT, D. S. *Análise Quantitativa na Concessão de Crédito versus Inadimplência: um estudo empírico*. Porto Alegre: UFRGS, 2003. Dissertação (Mestrado em Administração), Programa de Pós-Graduação em Administração, Universidade Federal do Rio Grande do Sul, 2003.

FARAGGI, D.; SIMON, R. The maximum likelihood neural network as a statistical classification model. *Journal of Statistical Planning and Inference*, v.46, n.1, p.93-105, jul. 1995.

FEELDERS, A. J. Credit Scoring and Reject Inference with Mixture Models, *International Journal of Intelligent Systems in Accounting Finance and Management*, v.9, n.1, p.1-8, mar. 2000.

GOUVÊA, M. A.; GONÇALVES, E. B. Análise de Risco de Crédito com o Uso de Modelos de Redes Neurais e Algoritmos Genéticos. In: *IX SEMEAD – Seminários em Administração FEA-USP*, 2006, São Paulo. Anais.

GUIMARÃES, I. A.; CHAVES NETO, A. Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor. *RAE Eletrônica EAESP/FGV*, São Paulo, v.1, n.1, jul.-dez. 2002.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. *Análise multivariada de dados*. 5.ed. Porto Alegre: Bookman, 2005.

HAND, D. J.; HENLEY, W. E. Statistical Classification in Consumer Credit Scoring: a Review. *Journal of the Royal Statistical Society A*, v.160, n.3, p.523-541, 1997.

HAYKIN, S. *Redes neurais: princípios e prática*. Trad. Paulo Martins Engel. 2.ed. Porto Alegre: Bookman, 2001.

HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. New York: John Wiley & Sons, 1989.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 5.ed. Upper Saddle River: Prentice Hall, 2002.

KASSAI, J. R.; KASSAI, S. O termômetro de Kanitz. In: *XXII ENANPAD – Encontro da Associação Nacional dos Programas de Pós-graduação em Administração*, 1998, Foz do Iguaçu. Anais.

KIMURA, H.; PERERA, L. C. J.; DONZELLI, C. R.; SILVA FILHO, A. C.; LIMA, F. G. Aplicação de Redes Neurais na Análise e na Concessão de Crédito ao Consumidor. In: *XXIX ENANPAD – Encontro da Associação Nacional dos Programas de Pós-graduação em Administração*, 2005, Brasília. Anais.

LAWRENCE, D. *Handbook of Consumer Lending*. New Jersey: Prentice Hall. 1992

LEWIS, E. M. *An Introduction to Credit Scoring*. San Rafael: Fair, Isaac and Co., Inc. 1992.

MARQUES, J. M.; LIMA, J. D. A Estatística Multivariada na Análise Econômico-Financeira de Empresas. *Revista FAE*, Curitiba, v.5, n.3, p.51-59, set.-dez. 2002.

McALLISTER, P. H.; MINGO, J. J. Commercial loan risk management, credit-scoring and pricing: The need for a new shared data base. *Journal of Commercial Bank Lending*, v.76, n.9, p.6-22, may 1994.

MENDES FILHO, E. F.; CARVALHO, A. C. P. L. F.; MATIAS, A. B. Utilização de redes neurais artificiais na análise de risco de crédito a pessoas físicas. In: *III Simpósio Brasileiro de Redes Neurais*, 1996, Recife. Anais.

MESTER L. J. What's the Point of Credit Scoring? *Business Review - Federal Reserve Bank of Philadelphia*, p.3-16, sep.-oct. 1997.

OLIVEIRA, M. L. *Fatores Intervenientes na Capacidade de Atendimento de Praças de Pedágio*. Porto Alegre: UFRGS, 2004. Dissertação (Mestrado em Engenharia de Produção), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, 2004.

PEREIRA, A. P. F.; BARROSO, M. H.; NEPOMUCENO FILHO, F. Uso do Credit Score na Análise de Crédito de Pessoa Física. In: *Congresso Nacional de Excelência em Gestão*, nov. 2002, Niterói, RJ.

PEREIRA, S. L. G. Na mira do crédito. *GV Executivo*, v.5, n.1, p.31-36, fev.-abr. 2006.

PICININI, R.; OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. Mineração de Critério de Credit Scoring Utilizando Algoritmos Genéticos. In: *VI Simpósio Brasileiro de Automação Inteligente*, 2003, Bauru.

QUEIROZ, R. S. B. A importância dos modelos de Credit Scoring na concessão de crédito ao consumidor no varejo. In: *IX SEMEAD – Seminários em Administração FEA-USP*, 2006, São Paulo.

ROSENBERG, E.; GLEIT, A. Quantitative methods in credit management: A survey. *Operations Research*. v.42, n.4; p.589-613, jul.-aug. 1994.

SANTOS, J. *Análise de Crédito: Empresas e pessoas físicas*. São Paulo: Atlas, 2000.

SAUNDERS, A. *Medindo o risco de crédito: Novas abordagens para value at risk e outros paradigmas*. Rio de Janeiro: Qualitymark, 2000.

SCARPEL, R. A.; MILIONI, A. Z. Utilização conjunta de modelagem econométrica e otimização em decisões de concessão de crédito. *Pesquisa Operacional*, Rio de Janeiro, v.22, n.1, p.61-72, jan.-jun. 2002.

SCHRICKEL, W. K. *Análise de Crédito: concessão e gerência de empréstimos*. 3.ed. São Paulo: Atlas, 1997.

SCHUMACHER, M.; ROBNER, R.; VACH, W. Neural networks and logistic regression: Part I. *Computational Statistics and Data Analysis*, p.21, n.6, p.661-682, jun. 1996.

SECURATO, J. R. *Decisões financeiras em condições de risco*. 1.ed. São Paulo: Atlas, 1996.

SELAU, L. P. R. *Redes neurais artificiais no contexto estatístico*. Porto Alegre: UFRGS, 2000. Monografia apresentada para obtenção do grau de Bacharel em Estatística, Departamento de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, 2000.

SILVA, E. L.; MENEZES, E. M. *Metodologia da pesquisa e elaboração de dissertação*. 4.ed. rev. atual., Florianópolis: Laboratório de Ensino a Distância da UFSC, 2005.

SILVA, J. P. *Análise e Decisão de Crédito*. São Paulo: Atlas, 1988.

SILVA, J. P. *Gestão e análise de risco de crédito*. 4.ed. São Paulo: Atlas, 2003.

SILVA, J. P. Os dois lados do crédito. *GV Executivo*, v.5, n.3, p.68-72, jul.-ago. 2006.

SILVA, M. A. *Elaboração de um modelo de análise e concessão de créditos para pessoas físicas em um banco*. Florianópolis: UFSC, 2003. Dissertação (Mestrado em Engenharia da Produção), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 2003.

SMITH, M. *Neural Networks for Statistical Modeling*. International Thomson Computer Press, London, 235 p., 1996.

SOUSA, A. F.; CHAIA, A. J. Política de Crédito: Uma análise qualitativa dos processos em empresas. *Caderno de Pesquisas em Administração*, São Paulo, v.7, n.3, jul.-set. 2000.

STEINER, M. T. A.; CARNIERI, C.; KOPITKE, B. H.; STEINER NETO, P. J. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. *Revista de Administração*, São Paulo, v.34, n.3, jul.-set. 1999.

THOMAS, L. C. A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Consumers, *International Journal of Forecasting*, v.16, n.2, p.149-172, apr.-jun. 2000.

VASCONCELLOS, M. S. *Proposta de método para análise de concessões de crédito a pessoas físicas*. São Paulo: USP, 2002. Dissertação (Mestrado em Economia), Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2002.

VASCONCELLOS, R. S. *Modelos de Escoragem de Crédito Aplicados a Empréstimo Pessoal com Cheque*. Rio de Janeiro: FGV, 2004. Dissertação (Mestrado em Finanças e Economia Empresarial), Escola de Pós-Graduação em Economia, Fundação Getúlio Vargas, 2004.

ZÜGE, M.; CHAVES NETO, A. Análise do desempenho Empresarial por Meio de Métodos Estatísticos Multivariados. *Revista Paranaense de Desenvolvimento*, n.97, p.101-112, set.-dez. 1999.

APÊNDICE A - AGRUPAMENTO DE PROFISSÕES

Péssimo Desempenho	BABA COZINHEIRO PINTOR	PROMOTOR VENDAS ALMOXARIFE
Muito Mau Desempenho	AUX PRODUCAO CABELEIREIRO CONFEITEIRO GERENTE PADEIRO	PEDREIRO PORTEIRO RECEPCIONISTA VENDEDOR
Mau Desempenho	AUTONOMO AUX ADMINISTRATIVO AUX COZINHA AUX SERVICOS GERAIS COMERCIANTE	MANICURE MECANICO TEC ENFERMAGEM VIGILANTE
Desempenho Neutro	ATENDENTE COMERCIARIO DOMESTICA	INDUSTRIARIO MOTORISTA
Bom Desempenho	CAIXA DO LAR PENSIONISTA	SECRETARIA SERVENTE
Muito Bom Desempenho	AGRICULTOR BALCONISTA COSTUREIRO DIARISTA	OPERADOR METALUGICO AUX ENFERMAGEM
Excelente Desempenho	APOSENTADO	PROFESSOR

APÊNDICE B - AGRUPAMENTO DE CIDADES DE NASCIMENTO

Péssimo Desempenho	ALVORADA	
Muito Mau Desempenho	CRUZ ALTA ESTEIO PORTO ALEGRE	RIO GRANDE TRAMANDAI
Mau Desempenho	CANOAS GRAVATAI IJUI NOVO HAMBURGO PELOTAS	SAO BORJA SAO GABRIEL SAPIRANGA SAPUCAIA DO SUL URUGUAIANA
Desempenho Neutro	ALEGRETE CAMAQUA CANELA SANTANA DO LIVRAMENTO	SANTO ANGELO SAO FRANCISCO DE PAULA SAO LOURENCO DO SUL VIAMAO
Bom Desempenho	BAGE BUTIA CACAPAVA DO SUL GUAIBA MONTENEGRO OSORIO	PASSO FUNDO RIO PARDO SANTA CRUZ DO SUL SAO JERONIMO SAO LEOPOLDO SAO LUIZ GONZAGA
Muito Bom Desempenho	CACHOEIRA DO SUL CAXIAS DO SUL PALMEIRA DAS MISSOES SANTA MARIA	SANTA ROSA SANTA VITORIA DO PALMAR TAQUARA
Excelente Desempenho	CANGUCU ENCRUZILHADA DO SUL GIRUA HORIZONTALINA ROLANTE SANTO ANTONIO DA PATRULHA	SAO SEPE TAPES TORRES TRES DE MAIO TRIUNFO

APÊNDICE C - AGRUPAMENTO DE CEP RESIDENCIAL

	2 PRIMEIRAS POSIÇÕES	3 PRIMEIRAS POSIÇÕES	4 PRIMEIRAS POSIÇÕES
Péssimo Desempenho			9670 SÃO JERÔNIMO
Muito Mau Desempenho			9175 Aberta dos Morros - POA 9179 Restinga - POA 9191 Camaquã - POA 9192 Cavalhada/Camaquã - POA 9440 Águas Claras - VIAM 9449 NS Aparecida/Pq Índio Jari - VIAM 9481 Formozo/Passo Feijó - ALVO 9493 Dist Industrial/Cohab - CACH
Mau Desempenho		902 Farrapos/Navegantes/Humaitá - POA 906 Partenon/Jardim Botânico - POA 912 Protásio Alves/Rubem Berta - POA 915 Lomba do Pinheiro/Agronomia - POA 923 Mathias Velho/Harmonia - CANO 934 Lomba Grande/Santo Afonso - NH 941 GRAVATAÍ 945 Vila Augusta/Jd Universit. - VIAM	9117 Rubem Berta - POA 9172 Nonoai/Teresópolis - POA 9174 Cavalhada/Vila Nova - POA 9190 Tristeza/Vila Assunção - POA 9326 Centro/Vila Teópolis - EST 9329 Pq Primavera/Pq St Inácio - EST 9353 São Jorge/Vila Diehl - NH 9400 GRAVATAÍ 9441 Centro/Tarumã - VIAM 9442 Vila Elsa/Estalagem - VIAM 9444 Jd Krahe/St Onofre - VIAM 9482 Maria Regina/Sumaré - ALVO 9483 Tijuca/Piratini - ALVO 9485 Aparecida/Jd Algarve - ALVO 9490 Jardim América/Vila City - CACH 9607 Porto/Três Vendas - PEL 9618 CAMAQUÃ 9750 URUGUAIANA
Desempenho Neutro		908 Santa Tereza/Medianeira - POA 913 Vila Jardim/Vila Ipiranga - POA 914 Protásio Alves/Jardim Carvalho - POA	9332 Industrial/Ouro Branco - NH 9445 São Lucas/Florescente - VIAM 9447 St Cecília/Viamópolis - VIAM 9480 Maringá/Sumaré - ALVO 9494 Vila Vista Alegre - CACH 9495 Vila Bom Princípio/Pq Matriz - CACH
Bom Desempenho		922 Fátima/Rio Branco - CANO 924 Igara/São José/Guajuviras - CANO 925 GUAÍBA 930 SÃO LEOPOLDO 938 NOVA HARTZ, SAPIRANGA 955 OSÓRIO, CAPÃO DA CANOA 956 TAQUARA, CANELA, GRAMADO 957 BENTO GONÇALVES, GARIBALDI	9178 Lami/Belém Novo - POA 9328 Vila Esperança/Pq Amador - EST 9330 Centro - NH 9333 Liberdade/Ideal - NH 9334 Primavera/Petrópolis - NH 9354 Canudos/Mauá - NH 9443 Jardim Krahe/Sítio S.José - VIAM 9496 Pq Granja Esperança - CACH 9601 Centro - PEL 9617 SÃO LOURENÇO DO SUL 9674 ARROIO RATOS e CHARQUEADAS 9754 ALEGRETE
Muito Bom Desempenho	99 PASSO FUNDO	950 CAXIAS DO SUL 962 RIO GRANDE, STA VITÓRIA PALMAR 965 CACHOEIRA DO SUL, CAÇAPAVA 970 SANTA MARIA	9407 GRAVATAÍ
Excelente Desempenho	98 CRUZ ALTA	937 CAMPO BOM 958 ESTRELA, TAQUARI, VENÂNCIO AIRES 964 BAGÉ, DOM PEDRITO 966 RIO PARDO, PÂNTANO GRANDE 968 SANTA CRUZ DO SUL 971 SANTA MARIA, ITAARA 973 SÃO GABRIEL, LAVRAS DO SUL	9352 Guarani/Vila Nova - NH

APÊNDICE D - AGRUPAMENTO DE CEP COMERCIAL

	2 PRIMEIRAS POSIÇÕES	3 PRIMEIRAS POSIÇÕES	4 PRIMEIRAS POSIÇÕES
Péssimo Desempenho		912 Protásio Alves/Rubem Berta - POA	9670 SÃO JERÔNIMO
Muito Mau Desempenho		900 Centro/Farroupilha/Bom Fim - POA 906 Partenon/Jardim Botânico - POA 932 ESTEIO, SAPUCAIA DO SUL 948 ALVORADA	9174 Cavalhada/Vila Nova - POA 9190 Tristeza/Vila Assunção - POA 9192 Cavalhada/Camaquã - POA
Mau Desempenho		901 Azenha/Menino Deus/Praia Belas - POA 902 Farrapos/Navegantes/Humaitá - POA 908 Santa Tereza/Medianeira - POA 910 Passo D'Areia/Jardim Lindóia - POA 911 Sarandi/Rubem Berta - POA 913 Vila Jardim/Vila Ipiranga - POA 915 Lomba do Pinheiro/Agronomia - POA 933 Rio Branco/Primavera/Industrial - NH 940 GRAVATAÍ 949 CACHOEIRINHA 961 CAMAQUÃ, CAPÃO DO LEÃO	9175 Aberta dos Morros - POA 9179 Restinga - POA 9191 Camaquã - POA 9351 Centro/Hamburgo Velho - NH 9353 São Jorge/Vila Diehl - NH 9602 Fragata/Três Vendas - PEL
Desempenho Neutro		904 Auxiliadora/Petrópolis - POA	9380 SAPIRANGA
Bom Desempenho		905 São João/Floresta/Higienópolis - POA 925 GUAÍBA 934 Lomba Grande/Santo Afonso - NH 955 OSÓRIO, CAPÃO DA CANOA 956 TAQUARA, CANELA, GRAMADO 957 BENTO GONÇALVES, GARIBALDI	9178 Lami/Belém Novo - POA 9389 NOVA HARTZ 9441 Centro/Tarumã - VIAM 9601 Centro - PEL 9674 ARROIO RATOS e CHARQUEADAS
Muito Bom Desempenho	99 PASSO FUNDO	959 LAJEADO, ENCANTADO, PROGRESSO 962 RIO GRANDE, STA VITÓRIA PALMAR	
Excelente Desempenho	97 SANTA MARIA 98 CRUZ ALTA	937 CAMPO BOM 950 CAXIAS DO SUL 958 ESTRELA, TAQUARI, VENÂNCIO AIRES 964 BAGÉ, DOM PEDRITO 965 CACHOEIRA DO SUL, CAÇAPAVA 966 RIO PARDO, PÂNTANO GRANDE 968 SANTA CRUZ DO SUL	

APÊNDICE E – PESOS DOS NEURÔNIOS DA REDE NEURAL

Neurônios da camada oculta

	DCAD1	DCAD4	DCAD6	DCAD7	DCAD8	DCAD9	DSEXP	DPRM	DSPR	DCASADO	DTSERVET	DTSERVW	DPLHO	DRES_ALI	DCERCER12	DCERPER1	DCERPEOT	DCERCER6	DGRDFP1	DGRDFP2	DGRDFP3	DGRDFP4	DCENAT1	DCENAS1	DCENAT2	DCENAS2	DCENAT3	DCENAS3
N1	-2,322	-2,252	-4,002	-2,656	2,0474	1,1644	-2,504	-3,527	2,379	-1,2564	1,9402	2,6662	3,9942	2,2076	-2,0246	-4,0002	-4,0118	2,6726	1,9492	-2,648	-2,059	-3,7594	-3,9146	-2,6566	-3,3194	3,3562	-4,0006	
N2	-3,0646	-4,216	2,0642	-8,751	3,2992	-1,826	2,805	-1,1606	-2,0424	2,203	-3,1566	1,8194	-4,9822	-1,8998	-4,0002	-3,002	-4,4764	2,9022	-1,376	-4,0002	-1,4114	-2,384	-4,236	-3,6976	-1,9194	-2,998	-4,838	
N3	-3,9966	-2,41	-1,681	1,6852	3,5002	-1,0476	-3,2766	-3,3696	-2,4322	3,9918	2,163	2,6098	2,2062	2,9862	-3,3252	-4,0002	-4,5076	1,536	3,209	-3,0946	-1,8812	-2,0916	-2,4006	-3,999	-2,9532	2,8552	-3,993	
N4	-2,9292	2,9462	1,9722	2,5394	2,72	2,6272	-3,9746	-4,5004	-3,597	2,6276	2,7596	-3,9934	2,894	-4,0014	-2,9714	2,9992	-2,4556	3,919	1,9472	-1,9852	-3,9834	2,817	2,7542	-3,2002	-2,4112	-2,9512	-2,3606	
N5	-3,6464	2,3126	2,861	2,1596	3,8944	2,8086	-2,123	-1,4386	-3,1446	2,9252	2,996	1,8906	2,214	-1,7108	-4,766	-4,0002	-2,8522	1,9992	-1,3164	-3,2236	-2,3834	-2,335	-1,846	-3,9906	-4,0002	3,3296	-3,9922	
N6	-3,4756	2,719	1,5994	-8,7494	4	-1,258	-8,172	-1,4692	-1,7406	1,1904	-2,2546	1,6616	-4,4964	-1,0044	-4,0002	-2,7306	-4,3974	2,4744	-1,215	-2,986	-2,0466	-2,3036	-4,414	-1,9624	-2,9482	-2,9666	-1,2664	
N7	-1,4448	-4,0002	-3,9992	-4,001	-4,2094	-3,997	-3,9994	-3,9992	-3,676	-3,9992	-3,9994	2,8702	-3,9992	2,2014	-3,999	-4	-3,9992	-4,001	-1,8996	-2,3136	-3,83	-4,0002	-4,0382	-3,9984	-4,0002	-4,0004	-3,9986	
N8	-3,6352	-1,5474	-1,2434	1,5054	3,139	-8,31	2,7776	-1,411	2,6896	3,3634	3,2474	2,238	-1,6172	1,0776	2,453	2,7934	2,3256	1,9382	-4,0626	-1,0512	-4,7902	-3,0706	-2,2292	-3,5816	2,467	3,9722	-3,5982	
N9	-4	2,851	2,6722	-3,4236	3,6322	1,8644	-4,3934	-6,9746	-5,2176	-8,103	-5,2416	3,1046	-2,5604	-6,3074	-3,9986	-3,391	-7,8176	2,6306	-3,6142	-3,2236	-3,591	-5,4444	-3,4272	-3,9952	-3,9984	-4,0002	-4,0004	-3,9986
N10	3,1862	-3,1764	-2,4036	-1,6732	-4,4234	-2,3734	-1,9422	1,5262	-1,875	2,0226	-8,891	2,6642	1,4716	2,115	1,2852	-8,462	-8,9992	-3,9996	2,5772	-1,9524	2,037	2,9442	-2,633	-1,6224	-3,0666	-3,9826	-3,9916	
N11	-2,5442	-6,1672	-1,4632	-1,301	2,598	1,5762	-1,8894	-2,1894	-2,6212	3,8312	1,8212	2,1452	2,4004	-8,8182	-1,9322	-3,8894	-1,1864	2,111	1,0444	-4,687	-2,9032	-3,4476	-2,2276	-2,8994	-2,282	1,83	-1,8016	
N12	-4,9596	2,89	-2,2632	-3,3062	-2,86	1,0119	2,5796	-4,0916	-3,2706	-3,5194	-2,556	-3,9976	2,437	1,484	1,3594	1,2894	-2,046	2,2546	-4,7396	-3,449	2,5094	2,5112	2,614	2,512	2,2406	-3,3796	-3,9994	
N13	-2,6212	-1,0232	2,9422	2,4666	2,04	-3,3256	2,8916	-1,6108	2,454	-4,9734	-3,3694	1,2854	-2,6232	2,017	-2,2296	-1,8094	-3,8996	-2,83	-1,4812	-3,981	1,8854	2,9256	2,1524	-1,5946	-1,021	-1,136	-4,0226	
N14	-1,9536	2,0852	2,6752	1,9932	3,8964	-2,1766	-8,761	2,299	-4,7016	2,3512	2,0046	3,3036	-1,542	-3,9644	-3,2444	-4,328	-2,1804	3,8802	3,6942	1,5442	-2,689	3,5076	3,6436	-3,972	2,3206	1,1434	-3,5212	
N15	-2,118	-3,969	2,2412	3,7236	3,9026	-2,218	1,0894	-1,124	3,7294	3,9192	2,1162	-3,9632	-2,7062	-2,896	-1,7906	-2,4118	-1,1364	3,9806	-4,4306	-3,9942	-4,328	-1,7174	-3,9902	-3,942	-4,4608	2,5622	-1,5806	
N16	1,4716	-2,382	-8,671	-8,9494	-2,7924	-3,2692	2,1966	2,5104	2,0794	2,619	-3,9996	-3,992	-2,248	-3,9694	-2,2696	1,93	-2,095	2,2076	-1,8254	3,9696	-2,1142	2,8664	2,4164	-1,5792	-2,837	-1,5474	-2,8934	
N17	1,4656	-3,984	-2,2526	-2,397	2,917	-3,9106	-2,9674	-2,979	3,9546	-3,9446	2,6226	2,513	-2,914	2,811	-2,9992	-3,9446	-2,9662	1,2504	3,9994	2,9612	-2,0052	-3,9714	-2,9616	-2,0052	-3,9714	-2,9616	-2,0052	-3,9634
N18	-5,4514	-7,9994	-7,9994	-7,9994	-2,145	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	4,0002	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994	-7,9994
N19	-2,3816	2,0546	-1,856	-3,7398	3,9606	-1,1926	-3,6664	-3,9662	-3,9698	3,9606	3,0014	3,076	2,0598	1,185	-3,976	-3,9416	-3,948	2,407	3,9672	2,9136	-2,2904	-2,689	-3,956	-3,9644	-3,9894	2,6072	-2,6434	
N20	-4,8702	-3,0316	-2,093	-2,569	-2,0294	-3,304	-4,1954	-4,5312	-2,835	-1,0426	-4,0908	3,665	-1,0496	1,2922	-2,911	-2,0246	-3,9474	2,281	-1,6794	-1,5924	-3,2926	-3,4216	-1,1626	-2,863	-3,1154	-3,2662	1,0974	
N21	-1,8304	-4,4526	-1,47	1,101	-2,1496	2,3312	-4,0962	-8,1512	2,9256	-4,2472	2,937	-3,9894	2,5414	-3,9762	-4,723	2,0382	3,583	-2,9674	-3,9666	1,9444	-1,4572	1,404	-4,9802	1,444	2,835	-1,4676	-2,5952	
N22	-2,4022	-3,9994	-3,978	-3,978	2,89	1,5176	-3,9724	-3,9806	-1,553	-3,9504	3,3206	3,9906	-3,9276	-1,095	-3,963	-3,9494	-3,8316	3,9994	-4,215	3,9742	-3,866	-1,89	-4,1592	-3,9774	-3,9894	-1,922	-3,944	
N23	-1,727	2,4424	-1,445	2,5176	2,892	-2,737	-2,1024	1,3054	1,9856	-2,8542	-1,1422	-4,1284	-2,5734	-3,9894	-1,0324	-2,094	3,4602	-2,7408	1,0432	2,6582	1,589	2,009	-2,3212	-3,7056	2,8992	2,155	-3,9182	
N24	1,8816	-2,2406	-4,4796	-1,0242	-3,8864	2,5916	-2,6232	1,821	-3,667	2,8732	-2,2056	-1,8692	-1,672	3,9912	1,6744	1,665	-2,0206	2,0314	-2,6636	2,6896	2,8844	-2,7732	-2,8196	2,8102	3,9396	-2,1714	-3,9432	
N25	-2,9546	-2,6532	-2,1402	-4,1714	1,0576	-4,889	-3,7276	-3,627	-2,4882	3,9946	2,8134	2,4194	2,158	1,148	-3,9816	-4,0006	-4,0002	3,918	3,9936	-1,6982	-2,2224	1,9772	-3,3108	-3,997	-3,521	2,267	-3,9262	
N26	3,0524	2,205	3,0954	3,8176	-1,1162	-4,4646	2,4266	-1,1532	-2,7974	3,728	3,9546	2,3816	2,83	2,89	-4,1646	-3,7234	-1,049	2,7572	3,4136	2,633	-1,4776	-3,9876	-1,5824	-4,7396	-4,6606	2,84	-3,4244	
N27	-1,877	-1,5812	-1,6882	-3,9972	-4,0034	-2,9772	-3,992	-3,9954	-1,4516	-3,9852	2,8642	2,9004	-3,9762	-1,2004	-3,9634	-3,9636	-3,9552	-3,9822	-2,5486	2,0408	-3,994	-3,9982	-3,9926	-3,994	-3,9982	-3,9984	-3,9882	
N28	-2,8514	3,9944	-1,957	-1,5976	-8,9822	2,1096	2,1108	2,8706	3,89	-4,714	-3,9964	1,7602	-1,1256	-1,4512	2,9632	-1,7342	-2,7782	-3,4672	1,0922	2,3204	-2,8164	2,7598	2,6294	-2,8396	-2,6416	-4,0026	-3,9964	
N29	-1,1712	-2,606	-2,2654	2,187	4,0002	-4,1946	1,8234	-4,5462	-1,173	1,693	-2,2042	-4,036	1,7894	-4,0006	2,3214	-3,2462	-3,7682	1,237	-2,2992	-4,736	2,785	-3,0922	-2,5644	-2,6002	2,708	-2,7494	-2,301	
N30	2,347	-3,999	-3,2656	-4,0006	-3,8192	-3,6942	-3,9954	-4,0006	-3,6262	-3,9922	-1,9992	-2,8412	-3,9872	1,9694	-3,9954	-3,2462	-3,9008	-1,2254	2,6966	3,4732	-3,8696	-3,9996	-4,282	-3,995	-3,999	-2,4196	-3,995	
N31	-3,9214	2,0316	2,9646	1,0574	4,0002	2,8012	2,4006	-1,044	-1,1112	1,3712	-3,201	2,4974	-1,154	-1,4212	-4,0002	-1,2766	-7,3694	1,7444	-2,7802	-4,0002	-3,3872	-4,5014	-1,1672	-2,7134	-2,235	-5,368	-2,2442	
N32	-3,9274	1,9106	2,6894	2,0894	3,841	1,1066	2,3098	-8,8304	-4,491	1,472	-2,715	2,3736	-1,1796	-1,3296	-4,0002	-1,202	-7,2304	1,5034	-2,7882	-3,7326	-3,0366	-3,6324	-4,8866	-2,5576	-2,2354	-2,7084		
N33	-3,142	-2,7164	-2,0542	-3,991	-4,003	-2,293	-3,9884	-2,994	-2,9696	-3,9864	2,373	-3,971	2,3182	-3,9924	-3,8826	-4,0004	-2,986	-2,6916	-2,4532	-3,557	-3,9924	-3,9974	-3,9802	-3,996	-3,9974	-3,9802	-2,9974	
N34	-4,4614	-3,8054	-3,503	-3,7746	-3,9536	-2,878	-2,9674	-3,3232	-3,8146	-1,421	1,8674	3,771	-3,4536	1,358	-3,7904	-3,596	-3,8806	2,333	-2,0072	-4,8672	-3,8804	-3,996	-3,538	-1,6836	-2,782	-3,903	-2,4162	
N35	-2,9402	1,6962	2,2624	-2,147	3,9974	2,5984	2,334	-1,1782	-3,786	2,9108	-1,732	2,2344	-1,1296	-1,587	-4,0002	-1,1706	-7,3224	1,725	-2,2892	-4,0002	-2,8314	-3,9874	-4,613	-2,6896	-2,1342	-2,229	-2,6896	

Neurônios da camada de saída

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17		
tipo_cil	0,0408	0,0164	0,079	-1,2922	0,0564	0,006	0,1072	-1,042	-0,0114	-2,4544	0,0936	-2,089	-1,4422	1,9152	2,321	-1,586	1,599		
	N18	N19	N20	N21	N22	N23	N24	N25	N26	N27	N28	N29	N30	N31	N32	N33	N34	N35	BIAS
	-0,0324	1,876	0,0592	-2,114	1,3852	-1,987	-1,2264	0,1434	-0,9456	0,4626	-1,3094	-0,8252	0,3744	-0,0106	0,0146	0,5982	-0,1596	0,0024	2,3756

APÊNDICE F – RELATÓRIO DE ACOMPANHAMENTO DE PERFIL

ANÁLISE DEMOGRÁFICA							
LOJA: Nome da Loja XX				DATA FINAL: 30/06/2008			
CARACTERÍSTICA	PADRÃO	Mês da Escoragem (%)					
		jan/08	fev/08	mar/08	abr/08	mai/08	jun/08
Sexo	(%)	(%)	(%)	(%)	(%)	(%)	(%)
FEMININO	68,1						
MASCULINO	31,9						
QUI-QUADRADO	3,84	100,00	100,00	100,00	100,00	100,00	100,00
z2 crítico							
Idade	(%)	(%)	(%)	(%)	(%)	(%)	(%)
até 20	9,5						
21 a 25	12,5						
26 a 30	11,8						
31 a 35	10,8						
36 a 40	10,6						
41 A 50	18,4						
51 A 60	13,5						
+ DE 60	13,0						
QUI-QUADRADO	14,07	100,00	100,00	100,00	100,00	100,00	100,00
Estado Civil	(%)	(%)	(%)	(%)	(%)	(%)	(%)
CASADO	31,5						
CONCUBINADO	0,7						
DIVORCIADO	5,9						
OUTROS	6,2						
SEPARADO	2,4						
SOLTEIRO	45,2						
VIÚVO	8,1						
QUI-QUADRADO	12,59	100,00	100,00	100,00	100,00	100,00	100,00
Escolaridade	(%)	(%)	(%)	(%)	(%)	(%)	(%)
FUNDAMENTAL	49,5						
MÉDIO	42,0						
SUPERIOR	8,5						
QUI-QUADRADO	5,99	100,00	100,00	100,00	100,00	100,00	100,00
Tem filho?	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Não	66,3						
Sim	33,7						
QUI-QUADRADO	4,61	100,00	100,00	100,00	100,00	100,00	100,00
Tempo de Emprego	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Não Informado	73,8						
até 3 meses	1,6						
4 a 6 meses	2,1						
7 a 18 meses	7,0						
19 a 24 meses	2,1						
25 a 36 meses	3,1						
37 a 60 meses	3,5						
61 a 90 meses	2,3						
91 a 120 meses	1,2						
121 meses ou +	3,4						
QUI-QUADRADO	16,92	100,00	100,00	100,00	100,00	100,00	100,00

Página 1

Ocupação	(%)	(%)	(%)	(%)	(%)	(%)	(%)
APOSENTADO	19,4						
ASSALARIADO	37,4						
AUTÔNOMO	37,2						
FUNCIONÁRIO PÚBLICO	3,9						
PROFISSIONAL LIBERAL	2,1						
QUI-QUADRADO	9,49	100,00	100,00	100,00	100,00	100,00	100,00
Residência							
Residência	(%)	(%)	(%)	(%)	(%)	(%)	(%)
RESIDE COM PAIS	1,0						
ALUGADA	9,4						
CEDIDA	5,5						
OUTRAS	9,8						
PRÓPRIA	74,3						
QUI-QUADRADO	9,49	100,00	100,00	100,00	100,00	100,00	100,00
Profissão							
Profissão	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Não Informado	9,94						
Não Classificado	7,14						
Grupo de PROF. 1	1,2						
Grupo de PROF. 2	10,5						
Grupo de PROF. 3	27,3						
Grupo de PROF. 4	14,8						
Grupo de PROF. 5	8,6						
Grupo de PROF. 6	2,3						
Grupo de PROF. 7	18,3						
QUI-QUADRADO	15,51	100,00	100,00	100,00	100,00	100,00	100,00
CEP Residencial							
CEP Residencial	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Não Informado	0,1						
Não Classificado	7,70						
Grupo de CEP RES 1	0,7						
Grupo de CEP RES 2	14,7						
Grupo de CEP RES 3	27,9						
Grupo de CEP RES 4	11,3						
Grupo de CEP RES 5	23,8						
Grupo de CEP RES 6	5,1						
Grupo de CEP RES 7	8,6						
QUI-QUADRADO	15,51	100,00	100,00	100,00	100,00	100,00	100,00
CEP Comercial							
CEP Comercial	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Não Informado	45,4						
Não Classificado	10,36						
Grupo de CEP COM 1	0,9						
Grupo de CEP COM 2	8,7						
Grupo de CEP COM 3	12,5						
Grupo de CEP COM 4	3,8						
Grupo de CEP COM 5	9,1						
Grupo de CEP COM 6	2,1						
Grupo de CEP COM 7	7,2						
QUI-QUADRADO	15,51	100,00	100,00	100,00	100,00	100,00	100,00
Cidade Nascimento							
Cidade Nascimento	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Não Informado	8,18						
Não Classificado	14,98						
Grupo de CID NAT 1	1,7						
Grupo de CID NAT 2	26,9						
Grupo de CID NAT 3	11,2						
Grupo de CID NAT 4	12,1						
Grupo de CID NAT 5	9,5						
Grupo de CID NAT 6	7,9						
Grupo de CID NAT 7	7,5						
QUI-QUADRADO	15,51	100,00	100,00	100,00	100,00	100,00	100,00

Página 2

ANEXO A – FORMULAÇÃO DA RÉPLICA DA REDE NEURAL EM PLANILHA ELETRÔNICA

O procedimento consiste em reproduzir, em uma planilha eletrônica, o modelo obtido através das redes neurais para gerar os valores previstos. Para tal, o primeiro passo é identificar qual função de ativação foi utilizada para encontrar os valores de saída. No caso estudado, o *software* de modelagem de redes neurais foi configurado para utilizar a função sigmóide para encontrar os modelos.

Inicialmente são calculados os valores de saída entre a camada de entrada e a camada oculta e, utilizando-se desses valores, volta-se a usar a função sigmóide para achar os valores entre a camada oculta e a camada de saída. As equações 01 e 03 mostram as características da função sigmóide empregada para cada caso. Os valores a_{0j} são ajustes adicionais entre camadas de neurônios.

$$F(u_j) = \frac{1}{1 + e^{-u_j}} \quad (01)$$

onde, u_j é igual a:

$$u_j = a_{0j} + \sum_{i=1}^l a_{ij} x_i \quad (02)$$

onde

a_{ij} = pesos do nó de entrada i para o nó oculto j

x_i = valor de entrada do nó i

a_{0j} = valor de ajuste (bias) do nó oculto j

l = número de nós de entrada

$$F(v_k) = \frac{1}{1 + e^{-v_k}} \quad (03)$$

onde, v_k é igual a

$$v_k = b_{0k} + \sum_{j=1}^J b_{jk} y_j \quad (04)$$

onde

b_{jk} = pesos do nó oculto i para o nó de saída k

y_j = valor de saída para o nó oculto j

b_{0k} = valor de ajuste (bias) do nó de saída k

J = número de nós ocultos

De posse dos dados de entrada é possível, dispondo de uma rede previamente treinada, com seus respectivos valores de pesos e valores de ajuste, aplicar as equações 01 e 03. O *software* de redes neurais informa ao usuário os pesos e valores de ajuste quando solicitados. Com estes elementos, as equações podem ser facilmente inseridas em um ambiente de planilha eletrônica.

Os resultados de saída, obtidos com as equações na planilha eletrônica, devem ser confrontados com os resultados gerados pelo *software* em que a rede foi treinada. Esta verificação deve ser feita em todos os procedimentos aqui propostos. Se o procedimento foi realizado corretamente, o resultado de ambos deve ser idêntico.