



Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Matemática e Estatística

João Ulisses Barata da Silva

**USO DE TÉCNICAS DE CLASSIFICAÇÃO NA ANÁLISE DE
CONCESSÃO DE CRÉDITO**

Belém
2011

João Ulisses Barata da Silva

**USO DE TÉCNICAS DE CLASSIFICAÇÃO NA ANÁLISE DE
CONCESSÃO DE CRÉDITO**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática e Estatística –PPGME – para obtenção do título de Mestre em Estatística da Universidade do Federal do Pará.
Orientador: Prof. Dr. Joaquim Carlos Barbosa Queiroz

Belém
2011

Silva, João Ulisses B. da.

Uso de Técnicas de Classificação na Análise de Concessão de Crédito / (João Ulisses Barata da Silva); orientador, Joaquim Carlos Barbosa Queiroz – 2011.

57 folhas. il. 28cm

Dissertação (Mestrado) – Universidade Federal do Pará. Instituto de Ciências Exatas e Naturais. Programa de Pós-Graduação em Matemática e Estatística. Belém, 2011.

1. Credit Scoring. 2. Qualidade de Crédito. 3. Regressão Logística. 4. Lógica Fuzzy.

I. Queiroz, Joaquim Carlos Barbosa, orient. II. Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós – Graduação em Matemática e Estatística. III. Título.

Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Matemática e Estatística

João Ulisses Barata da Silva

**USO DE TÉCNICAS DE CLASSIFICAÇÃO NA ANÁLISE DE
CONCESSÃO DE CRÉDITO**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática e Estatística –PPGME – para obtenção do título de Mestre em Matemática e Estatística da Universidade do Federal do Pará.

Data da defesa: 25 de março de 2011

BANCA EXAMINADORA

Prof. Dr. Joaquim Carlos Barbosa Queiroz (Orientador)
Universidade Federal do Pará - PPGME

Prof. Dr. Marcus Pinto da Costa da Rocha (Membro)
Universidade Federal do Pará - PPGME

Prof. Dr. Sandro Dimy Barbosa Bitar (Membro)
Universidade Federal do Amazonas - ICEN

Prof. Dr. Carlos Tavares da Costa Júnior (Membro)
Universidade Federal do Pará - PPGEE

Belém
2011

*A minha família: Meus Pais, meu
irmão e minha noiva.*

Agradecimentos

Toda honra e toda glória ao Deus único e verdadeiro, por ter segurado firmemente em minhas mãos nas horas difíceis, fato este que me fez chegar até aqui.

Aos meus pais, João e Alzira, e meu irmão Marcos, amores da minha vida, força e o incentivo para vencer.

A minha Tia Lindalva e meu primo Junior pelo companheirismo e confiança, nos momentos de tristes e felizes, sempre presentes.

A futura esposa e o amor da minha vida, Nayane Caroline pelo grande apoio, companheirismo e confiança, sem você tudo seria mais difícil.

Ao Prof. Marcus Rocha pelo apoio e pela confiança em meu trabalho e pela oportunidade que me deu de usufruir de um ambiente que me proporcionou tanto crescimento acadêmico e profissional.

Ao meu orientador Prof. Joaquim Queiroz, pela calma na condução dos momentos aflitos no desenvolvimento desta dissertação, sua muito grato pela força.

Aos meus amigos, “*Família LAM*”, Prof^o Valcir, Prof^o Eduardo Brandão, Anderson Campelo e Edinaldo Alencar, Marcos Chagas, André Macias, Josiane Lira, pelos preciosos momentos de troca de conhecimento, mas também de experiências de vida.

A eterna e Grande Amiga para todos os momentos, Raquel Albuquerque.

A todos os professores que ainda que indiretamente, tenham contribuído com valiosos ensinamentos, em especial, Prof^a. Regina, Prof^o Heliton e Prof. Joaquim Queiroz.

Ao Prof^o, amigo e irmão Paulo Fernando pelos seus ensinamentos e atenção.

Ao Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará, por ter proporcionado esta grande oportunidade.

A FAPESPA, pela concessão da bolsa de mestrado e pelo apoio financeiro para a realização desta pesquisa.

"Deus de Aliança... Deus de Promessas... tudo pode Passar... tudo pode Mudar... mas tua Palavra vai se Cumprir...".

(Davi Sacer)

Resumo

Este trabalho consiste em uma sugestão de metodologia para análise de concessões de crédito, a partir de um estudo estatístico de informações sobre créditos concedidos num passado recente, tais como os hábitos de pagamento e variáveis cadastrais, financeiras e patrimoniais e de relacionamento com a instituição credora dos clientes analisados. Para alcançar os objetivos propostos, será utilizado o método estatístico de Regressão Logística integrado ao método de Inteligência artificial, Lógica Fuzzy. Para modelagem estudada, foi utilizada a base de dados de uma instituição financeira com registro de 5.917 usuários de cartão de crédito, com cadastro no período de jan./2000 a set./2003. Dessa base foi extraído o conhecimento através do modelo de Regressão logística, usando o método de seleção de variáveis *forward stepwise*, assim resultando em um modelo de análise de crédito com 9 variáveis significativas – *Grau de instrução, Sexo do cliente, Faixa de etária, Residência própria?, N° de dependente, Faixa Salarial, Quantidade de parcelas, Cheque devolvido?, Pendência no Serasa* e atendendo a todos os pressupostos de adequação do modelo, com a taxa de acerto de 82,6% (poder classificação). A partir do conhecimento adquirido pela ferramenta estatística regressão logística, foi elaborado um Sistema de Inferência Fuzzy do Tipo Mamdani, tendo como variáveis de resposta a *qualidade de crédito* e o *Risco de crédito*. Para melhor sensibilidade de resposta fuzzy, foram selecionadas 7 variáveis com melhor desempenho baseando-se na Razão de chance (OR) do modelo logístico – *Grau de instrução, Sexo do cliente, Idade, Salário, Quantidade de parcelas, Cheque devolvido, Pendência no Serasa* – Essas variáveis foram definidas como variáveis de entrada, suas saídas conseqüentes mostraram-se “sensíveis”, e correspondendo a proposta do trabalho. O modelo proposto neste trabalho apresentou uma forma interessante de simulação do conhecimento do especialista utilizando a Regressão Logística e Lógica Fuzzy. As taxas de acerto entre a Lógica Fuzzy e a Regressão Logística foram bem próximas uma da outra.

Palavras-Chave: Credit Scoring, Qualidade de Crédito e Regressão Logística e Lógica Fuzzy.

Abstract

This paper presents a suggested methodology for the analysis of extensions of credit, from a statistical analysis of information on loans in the recent past, such as variable payment habits and personal, financial and patrimonial and relationship with the lending institution customer analysis. To achieve the proposed objectives, we adopted a statistical method of logistic regression method integrated with artificial intelligence, fuzzy logic. Modeling study, we used the database of a financial institution with a record of 5,917 credit card users, register in the period jan./2000 set./2003. This knowledge base was extracted through the logistic regression model, using the method of forward stepwise variable selection, thus resulting in a credit analysis model with nine significant variables - degree of education, sex of client, age range, Housing own?, No. dependent, salary range, quantity of parts, Cheque returned?, Pendency in Serasa and given all the assumptions of model fit, with the hit rate of 82.6% (power rating). From the knowledge acquired by logistic regression statistical tool, was developed a Fuzzy Inference System Type Mamdani, having as response variables to credit quality and credit risk. For better response sensitivity fuzzy, 7 variables were selected with better performance based on the odds ratio (OR) of logistic model - Level of education, sex of client, Age, Income, Number of plots, Cheque returned Pendency in Serasa - These variables were defined as input variables, their resulting outputs proved to be "sensitive" and corresponding to the proposed work. The model proposed in this paper has presented an interesting way to simulate the expert knowledge using Logistic Regression and Fuzzy Logic. Hit rates between the fuzzy logic and logistic regression were very close to each other.

Keywords: Credit Scoring, Credit Quality and Logistic Regression and Fuzzy Logic.

Lista de Figuras

2.1	Esquema de modelo de mensuração de risco	6
2.2	Ciclo de desenvolvimento de um modelo	11
3.1	Curva logística	15
3.2	Estrutura de um controlador de Lógica <i>Fuzzy</i>	27
3.3	Função de Pertinência Triangular	30
3.4	Função de Pertinência Trapezoidal	31
3.5	Função de Pertinência Gaussiana	31
3.6	Conjunto Fuzzy A e o seu Complemento A	32
3.7	Interseção entre conjuntos <i>fuzzy</i>	33
3.8	União entre conjuntos <i>fuzzy</i>	34
3.9	Exemplo de variável lingüística	37
5.2.1	Distribuição do Estado Civil do Cliente quanto a Qualidade de Crédito	46
5.2.2	Distribuição do Grau de Escolaridade do Cliente quanto a Qualidade de Crédito ...	47
5.2.3	Distribuição da situação de residência do cliente quanto à Qualidade de Crédito ...	48
5.2.4	Distribuição da Restrição do cliente no Serasa quanto à Qualidade de Crédito	48
5.4.1	Resumo da Ferramenta do Sistema Fuzzy	55
5.4.2	Função de pertinência associada a variável de saída “ <i>Qualidade de Crédito</i> ”	57
5.4.3	Função de pertinência associada a variável de saída “ <i>Risco de Crédito</i> ”	57
5.4.4	Funcionamento do Modelo Fuzzy para Análise de Crédito	58

Lista de Tabelas

3.1	Conjuntos tradicionais	28
3.2	Propriedades Fundamentais das operações sobre Conjuntos Tradicionais	29
3.3	Operações e Relações com Conjuntos Fuzzy	32
3.4	Métodos de Defuzzificação	40
5.2.1	Distribuição dos Clientes quanto a Qualidade de Crédito	45
5.2.2	Distribuição do Gênero do Cliente quanto à Qualidade de Crédito	46
5.2.3	Distribuição do regime de casamento do cliente quanto à Qualidade de Crédito ...	47
5.2.4	Resumo Estatístico do Salário Mensal (em R\$), quanto à qualidade crédito dos Clientes	49
5.2.5	Resumo Estatístico da Idade, quanto à qualidade crédito dos Clientes	49
5.2.6	Resumo Estatístico do Valor do Contrato (em R\$), quanto à qualidade crédito dos Clientes	50
5.2.7	Resumo Estatístico do Número de Parcelas, quanto à qualidade crédito dos Clientes	50
5.3.1	Recategorização das variável “ <i>Grau de instrução</i> ”	51
5.3.2	Variáveis e Coeficientes Estimados no Modelo Ajustado	52
5.3.3	<i>Teste de Bondade de ajuste de Pearson, Deviance e Hosmer-Lemeshow</i>	53
5.3.4	Taxa de acerto do modelo logístico da Amostra de Treinamento	54
5.3.5	Taxa de acerto do modelo logístico da Amostra de Validação	55
5.4.1	Resumo dos Parâmetros do Sistema fuzzy	56
5.4.2	Variáveis lingüísticas do Sistema Fuzzy	56

Sumário

RESUMO	ixii
ABSTRACT	x
LISTA DE FIGURAS	xxii
LISTA DE TABELAS	xiv
1. INTRODUÇÃO	1
1.1 ASPECTOS GERAIS.....	1
1.2 JUSTIFICATIVA E IMPORTÂNCIA DO TRABALHO.....	2
1.3 OBJETIVO.....	3
1.3.1 Objetivo Geral.....	3
1.3.2 Objetivos Específicos.....	3
1.4 ESTRUTURA DO TRABALHO.....	3
2. RISCO DE CRÉDITO	4
2.1 CRÉDITO.....	4
2.1.1 A Evolução Histórica do Crédito.....	5
2.1.2 Crédito ao Consumidor.....	6
2.1.3 Análise e Concessão de Crédito.....	7
2.1.4 Risco de Crédito.....	8
2.2 CREDIT SCORING.....	10
3. MÉTODOS	13
3.1 ANÁLISE DE REGRESSÃO.....	13
3.1.1. Regressão Logística.....	14
3.1.1.2 Histórico.....	15
3.1.1.3 Descrição do Modelo.....	16
3.1.1.3.1 Estimação do Parâmetro.....	18
3.1.1.3.2 Função de Verossimilhança.....	18
3.1.1.3.3 Método de Escolha das Variáveis Explicativas do Modelo.....	19
3.1.1.4 Razão de Chance.....	21
3.1.1.5 Diagnostico do Modelo Logístico.....	22
3.1.1.6 Teste Wald de Bondade de Ajuste.....	23
3.1.1.6.1 Função de Deviance de Bondade de Ajuste.....	23
3.1.1.6.2 Gráfico do Delta Qui - quadrado.....	24

3.1.1.7 A Escolha do Score de Corte.....	24
3.2 LÓGICA FUZZY.....	25
3.2.1 Considerações Gerais	25
3.2.2 Descrição da Lógica Fuzzy.....	26
3.2.3 Funções de Pertinência Fuzzy mais Comuns.....	27
3.2.4 Fuzzificadores e Defuzzificadores.....	30
3.2.4.1 Fuzzificadores.....	31
3.2.4.1.1. Tipos de Fuzzificadores.....	31
3.2.4.2. Defuzzificadores	32
3.2.4.2.1 Tipos de Defuzzificadores.....	32
3.2.5.4.1 Conclusão.....	33
4. DESCRIÇÃO DO ESTUDO.....	34
4.1 ASPECTOS GERAIS.....	34
4.2 POPULAÇÃO E AMOSTRA.....	34
4.3 DEFINIÇÃO DAS VARIÁVEIS.....	35
4.4 METODOLOGIA.....	36
5. APLICAÇÃO.....	38
5.1 ASPECTOS GERAIS.....	38
5.2 ANÁLISE EXPLORATÓRIA DOS DADOS	38
5.3 REGRESSÃO LOGÍSTICA.....	43
5.3.1. Construção do Modelo Logístico	43
5.3.2. Estimação dos Parâmetros	45
5.3.2.1. Teste de significância de cada variável	46
5.3.1.2. Teste de significância do modelo – Teste de bondade de ajuste	46
5.2.1.3. Teste G	46
5.2.1.4. Análise da Razão de chance	47
5.2.1.5 Ponto de corte e Validação do modelo ajustado	47
5.4. LÓGICA FUZZY	48
6. CONSIDERAÇÕES FINAIS E PERSPECTIVAS	55
REFERENCIAS BIBLIOGRÁFICAS	57

Capítulo 1

INTRODUÇÃO

1.1 ASPECTOS GERAIS

O termo crédito, no sentido restrito, consiste na entrega de um bem ou de um valor monetário mediante uma promessa de pagamento em uma data futura. O risco de crédito se origina na possibilidade dessa promessa não ser cumprida por parte do tomador de crédito (SILVA, 1998). A concessão de crédito tem um papel fundamental na economia do país. No Brasil a indústria do crédito é relativamente pequena em relação à dos países desenvolvidos, mas a liberação de crédito ao consumidor no Brasil vem apresentando altas taxas de crescimento após a implantação do Plano Real e a estabilização da inflação.

Há alguns anos atrás, para o processo para fazer uma solicitação de crédito, o cliente preenchia uma proposta que seria avaliada por um ou mais analistas que apresentavam um parecer em relação ao pedido (SEMOLINI, 2002). Apesar de eficaz, este processo era lento, por não permitir a análise de muitos pedidos. Com isso as instituições financeiras – empresas financiadoras de crédito – necessitam da elaboração de uma análise técnica previamente à decisão de conceder ou não o capital ao proponente, com o objetivo de acelerar a avaliação das propostas.

O risco de concessão pode ser calculado com base em informações e dados históricos fundamentados, para que a decisão seja tomada a partir de estimativas julgadas aceitáveis. Deste modo, a análise de crédito procura fazer avaliação adequada dos riscos envolvidos, a partir de dados e informações fidedignas sobre o tomador de recursos e o negócio pretendido, de modo a reduzir a probabilidade de perda. Logo, a avaliação de risco do cliente tem por finalidade básica calcular o perfil de risco de crédito, definindo a probabilidade de inadimplência no segmento a que ele pertence.

Existem no mercado diversos métodos e modelos de classificação de risco, sendo que cada um oferece maior ou menor grau de confiabilidade, dependendo do critério e do rigor

definido na classificação, mas uma coisa deve ter em vista, independente do método de avaliação de risco escolhido, sempre existirá um paradoxo: quanto mais rigoroso ele for, maior a quantidade de clientes “bons” que irão ficar fora da carteira; quanto mais flexível, maior o número de clientes “ruins” que serão atendidos. De uma forma mais prática, observa-se que quanto menores forem as malhas da peneira de seleção, maior a quantidade de clientes bons que serão reprovados na análise, e ao contrário, quanto maiores forem às malhas da peneira, maior a quantidade de clientes ruins que poderão ser aprovados.

Os modelos representam o conhecimento adquirido e a experimentação humana que, somados, podem ser aplicados à explicação da maneira como as pessoas se comportam ou como as coisas funcionam, facilitando a compreensão de um fenômeno e, eventualmente, sua exploração.

Com isso, pretende-se descobrir as variáveis ou a combinação delas que influenciam ou agravam o risco do cliente, ponderando a segurança, manutenção, elevação ou redução do limite de crédito do cliente. Para alcançar os objetivos propostos, será utilizado o método estatístico de Regressão Logística (Hosmer & Lemeshow, 1989) integrado ao método de Inteligência artificial, Lógica Fuzzy (Kartalopoulos, 1996), estes métodos terão por base o histórico de concessões realizadas no passado recente. Cada um dos métodos irá gerar um *score* ao “proponente” – candidato a obter o crédito – a partir de informações consideradas relevantes na definição da sua capacidade de pagamento do crédito. Esse sistema de análise deve ser desenhado de forma que a chance de um mau pagador ter um “*score alto*” seja pequena (pequena possibilidade de aprovação do pedido de crédito a um mau pagador), e também de forma que a chance de um bom pagador ter um *score baixo* seja pequena (pequena possibilidade de recusa do pedido de crédito a um bom pagador).

Espera-se, deste modo, contribuir no setor de tecnologia da informação e comunicação com a implementação de metodologias que permitam a análise e avaliação de grandes bases de dados para planejamentos e auxílio em tomadas de decisão relacionadas a riscos de créditos em instituições financeiras.

1.2 JUSTIFICATIVA E IMPORTÂNCIA DO TRABALHO

Nos últimos anos vem crescendo a inadimplência no cenário econômico e comercial, com isso cresce o interesse das organizações financeiras procurando técnicas estatísticas eficientes, que por meio de probabilidades associadas aos clientes

os classifique como um “*cliente bom*” ou um “*cliente ruim*”, com base em seus dados cadastrais e seu histórico recente de operações financeiras, minimizando assim o risco de falha na concessão do crédito.

Na literatura pesquisada, principalmente no Brasil, encontram-se poucos estudos que abordam lógica fuzzy como ferramenta para construção de modelos de *credit scoring* (*técnica de classificação*). Em contrapartida, a regressão logística é uma técnica largamente empregada neste tipo de problema. Por esta razão, julgou-se oportuno utilizar as duas técnicas conjuntamente para análise questão.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Desenvolver uma metodologia para análise de crédito, usando como suporte as técnicas de classificação “Regressão Logística e Lógica Fuzzy”, de forma a assegurar um processo de crédito seguro e de acordo com as necessidades operacionais do segmento.

1.3.2 Objetivos Específicos

- Selecionar as variáveis a serem utilizadas em cada uma das técnicas;
- Definir critérios para aferir o poder de discriminação das variáveis;
- Desenvolver um modelo de regressão logística;
- Desenvolver um modelo linguístico a partir da teoria dos conjuntos fuzzy;
- Identificar as variáveis com maior poder de discriminação dos clientes catalogados nos grupos dos “*Bons*” e “*Maus*” clientes.

1.4 ESTRUTURA DO TRABALHO

Este trabalho é composto de seis capítulos, onde os conteúdos destes capítulos estão resumidos abaixo.

Este *Capítulo 1* apresentou as principais considerações sobre o trabalho, como os Aspectos Gerais, a Importância, Objetivos e Estrutura do Trabalho;

O *Capítulo 2* irá mostrar a fundamentação teórica, contendo conceitos de crédito, risco e modelos de *credit scoring*;

O *Capítulo 3* permite uma visão geral das técnicas adotadas no estudo;

O *Capítulo 4* é descrita as particularidades deste estudo, como a explicação do problema estudado;

O *Capítulo 5* ilustra uma visão mais detalhadas das técnicas e a forma como elas foram adotadas;

O *Capítulo 6* apresenta as conclusões e sugestões para trabalhos futuros.

Capítulo 2

Risco de Crédito

“O valor de nossas expectativas sempre significa algo entre o melhor que podemos esperar e o pior que podemos temer” - Jacob Bernoulli, 1654-1705. Mesmo com uma simples definição de conceito de risco o assunto é complexo. A etimologia da palavra risco deriva do italiano antigo “rischio”, cujo significado é ousar.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. CRÉDITO

Como a palavra já diz, crédito é confiança. Confiança em uma pessoa que hoje se compromete a cumprir uma obrigação futura. As obrigações de cunho pecuniário, por meio do crédito, agilizam as atividades econômicas principalmente pelo fato de uma empresa satisfazer hoje a uma necessidade, pagando o seu preço no futuro (FILHO, 1990).

Assaf e Tibúrcio (1999, p.99) definem que “*Crédito diz respeito à troca de bens presentes por bens futuros. De um lado uma empresa que concede crédito troca produtos por uma promessa de pagamento futuro. Já uma empresa que obtém crédito recebe produtos e assume o compromisso de efetuar pagamentos futuros*”.

Encontra-se em Silva (1998, p.67) definição sobre o significado crédito: “*Crédito (Do lat.creditu.) S.m. 1. Segurança de que alguma coisa verdadeira; confiança: Suas afirmações merecem crédito. A crédito. Recebendo o objeto sem pagar no ato de comprar, ou entregando-o sem receber no ato de pagamento; fiado: comprar a crédito, vende a crédito. Levar a crédito. Creditar*”

Outra definição de crédito pode ser vista em Schrickel (2000, p.6) “*Crédito é todo ato de vontade ou disposição de alguém de destacar ou ceder, temporariamente, parte de seu patrimônio a um terceiro, com a expectativa de que esta parcela volte a sua posse integralmente, após decorrido o tempo estipulado*”.

O crédito pode fazer com que as empresas aumentem seu nível de atividade; estimular o consumo influenciado na demanda; cumprir uma função social ajudando as pessoas a obterem

moradia, bens e até alimentos; facilitar a execução de projetos para os quais as empresas não disponham de recursos próprios suficientes. A tudo isso, por outro lado, deve-se acrescentar que o crédito pode tornar empresas ou pessoas físicas altamente endividadas, assim como pode ser parte componente de um processo inflacionário (SILVA, 1998).

Nos dias atuais, crédito é um negócio essencial. “O maior desafio desta indústria é torna o crédito largamente disponível; assim tantas as pessoas quanto possíveis terão a oportunidade de utilizar uma poderosa ferramenta” (LEWIS, 1992, p.2). Entretanto, tornar o crédito largamente disponível não significa distribuir crédito indistintamente para todos que o solicitam; existe um fator associado ao crédito ao consumidor que é decisivo na decisão de disponibilizar ou não crédito: risco.

2.1.1. Evolução histórica do crédito

As instituições financeiras surgiram inicialmente em Roma a partir dos cambistas que se aproveitavam da diversidade de moedas existentes na época para realizar trocas entre elas, obtendo sempre pequenas vantagens em suas negociações.

Os cambistas normalmente ficavam em pequenos bancos em lugares de movimento, tais como igrejas, estabelecimentos públicos, praças, etc., daí a origem do nome *banqueiros*. Posteriormente essas atividades foram expandidas pela possibilidade de recebimento de depósitos em dinheiro e a oferta de empréstimos com recursos próprios mediante a cobrança de um acréscimo no valor futuro.

Segundo Caouette, Altman e Narayanan (1998), os primeiros banqueiros na Europa Medieval frequentemente cobravam dos clientes pequenas tarifas em função dos custos associados a guarda de seus recursos. Contudo, não demorou muito para que eles percebessem que, emprestando esses recursos a outros, poderiam fazer dessa atividade um negócio rentável.

De acordo com Perera (1998), durante a Idade Média a doutrina cristã não incentivou as operações de crédito por considerar que os juros sobre os empréstimos eram proibidos, contudo, é importante destacar a distinção entre juros e usura. Para a igreja a proibição era contra a usura, na qual é cobrado mais do que lhe foi concedido.

Apesar da igreja manter sempre oposição à cobrança de juros em empréstimos no início do processo de expansão marítima a restrição foi reduzida, pois os detentores de capital desejavam participar do resultado dos empreendimentos financiados. Com a evolução da

sociedade de consumo essa restrição foi perdendo força sendo inclusive aceito pela igreja, em determinadas situações, as operações de crédito.

Com a evolução dos créditos bancários começou a surgir o chamado crédito comercial, que era negociado entre as empresas. Inicialmente as operações tinham como principal função a ajuda mútua entre comerciantes de forma a suprir faltas momentâneas de mercadorias. Contudo, essas operações foram, pouco a pouco, se tornando mercantis, ou seja, com objetivo de obtenção de lucro.

No início do crescimento do crédito comercial, as operações de vendas financiadas eram concedidas principalmente aos comerciantes localizados na mesma região. Todavia, a necessidade de expansão comercial para as outras localidades fez com que os comerciantes começassem a oferecer financiamentos aos compradores das suas mercadorias. Como forma de evitar riscos pela venda a prazo entre localidades surgiu a chamada Letra Cambial, que representava um título de reconhecimento de dívida pelo comprador de mercadorias. Outra vantagem das letras cambiais é a possibilidade de recebimento antecipado, por parte do credor, dos seus recursos através da venda desses títulos. As operações de recebimento antecipado das letras cambiais apresentam grande semelhança ao desconto bancário atual.

Na atualidade, as operações de créditos comerciais não se restringem apenas às transações entre as empresas, havendo também o chamado crédito ao consumidor. Perera (1998) observa que, apesar de não se conhecerem dados precisos sobre o início e o desenvolvimento do crédito ao consumidor, esse tipo de operação sempre existiu como forma de facilitar a efetivação da venda de mercadorias.

2.1.2 Crédito ao Consumidor

A expressão crédito ao consumidor pode ser entendida como uma forma de comércio onde uma pessoa física obtém dinheiro, bens e serviços e compromete-se a pagar por isso futuramente, acrescendo ao valor original um prêmio (juros) (SANTOS, 2000).

Atualmente, o crédito ao consumidor é uma grande indústria que opera no mundo. Grandes varejistas impulsionam suas vendas, fornecendo crédito. Empresas automobilísticas, bancos e outros segmentos utilizam as linhas de créditos para obter lucros. Por outro lado, o crédito ao consumidor injeta recursos na economia, permitindo a produção e a expansão econômica de um país, trazendo desenvolvimento à nação (LEWIS, 1992).

No entanto, muitos consumidores ainda não sabem como utilizar corretamente os créditos disponíveis no mercado e acabam tornando-se inadimplentes. As instituições financeiras,

preocupadas com esta realidade, passam a investir em iniciativas a fim de minimizar os seus riscos, tornando a análise de crédito uma ferramenta crucial no processo de concessão de crédito ao consumidor.

2.1.3 Análise e Concessão de Crédito

Análise de Crédito é o momento no qual o *agente cedente* irá avaliar o potencial de retorno do tomador do crédito, bem como, os riscos inerentes à concessão. Tal procedimento é realizado, também, com o objetivo de identificar os clientes que futuramente poderão não honrar com suas obrigações, acarretando uma situação de risco de caixa à Instituição *cedente*.

Esta avaliação de crédito constitui-se por um processo organizado para analisar dados, de maneira a possibilitar o levantamento das questões certas acerca do tomador do crédito. "Este processo cobre uma estrutura mais ampla do que simplesmente analisar o crédito de um cliente e dados financeiros para a tomada de decisão com propósitos creditícios" (BLATT, 1999, p.93).

A concessão de crédito tem um papel fundamental na economia de um país. Cerca de dois terços do produto interno bruto (PIB) dos Estados Unidos, por exemplo, decorre do consumo¹. No Brasil, a indústria de crédito é, relativamente ao tamanho da economia, bem menor do que os países desenvolvidos². Porém, o crédito ao consumidor vem apresentando altas taxas de crescimento após a implantação do Plano Real e o controle da inflação³.

As instituições que concedem crédito necessitam de um procedimento para decidir se emprestarão ou não capital a um proponente. Essa decisão é fundamental para o resultado financeiro da instituição, já que o lucro dos credores está diretamente associado à proporção de candidatos aprovados e ao percentual de clientes que pagam as dívidas contraídas.

Segundo Santos (2000), o processo de análise e concessão de crédito recorre ao uso de duas técnicas: a técnica subjetiva e a técnica objetiva ou estatística. A primeira diz respeito à técnica baseada no julgamento humano e a segunda é baseada em processos estatísticos, matemáticos onde a precisão é mais ampla. Em relação à primeira técnica, Schrickel (2000, p.27) observa que: "a análise de crédito envolve a habilidade de fazer uma decisão de crédito, dentro de um cenário de incertezas e constantes mutações e informações incompletas".

¹ Depois da guerra, confiança cai nos Estados Unidos. Folha de São Paulo, 14 jun.2003. Caderno Dinheiro, p.B8.

² Explosão de crédito é receita para problema, diz Moody's. Folha de São Paulo, 23 nov.2003. Caderno Dinheiro, p.B4.

³ BCB – Séries Temporais. Banco Central do Brasil. <http://www4.bcb.gov.br/pec/series/port>. Acesso em 10 de maio de 2009.

Para Schrickel (2000), as técnicas específicas 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.

2.1.4 Risco de Crédito

O risco de crédito é uma das formas mais antigas de risco no mercado financeiro, ou seja, é consequência de uma transação financeira contratada entre o fornecedor (doador de crédito) e um usuário (tomador de crédito). Segundo Jorion (1997), este risco pode ser definido como sendo a possibilidade da contraparte não cumprir as suas obrigações monetárias contratuais relativas às transações financeiras. Esse não cumprimento das obrigações contratuais é chamado de **inadimplência**.

O risco de inadimplência constitui a principal variável para a modelagem de risco de crédito, pois este auxilia na constituição de provisões, na precificação das operações de crédito e no estabelecimento de limites de crédito (AKIAMA, 2008).

Toda operação de crédito apresenta inadimplência esperada. Por isso a mensuração de risco de crédito tem por exigência um conhecimento prévio da probabilidade de inadimplência associada à classificação de risco, permitindo assim minimizar os custos dessa inadimplência no preço de cada operação. Para Caouette, Atman e Narayanan (1998), o processo de avaliação da capacidade financeira do cliente pode se comparado como a de um alfaiate, ou seja, feita sobre medida para as características do comprador.

Bessis (1998, p.81) acrescenta que “Risco de crédito é definido pela perda no evento de não pagamento do devedor, ou no evento de não deterioração da qualidade do crédito do devedor”⁴. Sob esse entendimento, a elevação do risco é resultante não somente pela falta de pagamento de uma obrigação, mas também pela redução da capacidade de pagamento do devedor.

Atualmente as instituições financeiras perceberam que as razões para a sua existência é a administração de risco. Por isso, buscaram o que isto significava e como poderiam transformar o risco em algo plenamente administrável.

⁴ Tradução de “*Credit risk is defined by the losses in the event of default of the borrower, or in the event of deterioration of the borrower’s credit quality.*”

Para melhor entender o risco de crédito, necessário verificar o processo decisório. Este processo incorpora a obtenção de um grande número de informações dos clientes, interessados em uma decisão de crédito.

Em geral os modelos de mensuração de risco de crédito utilizam variáveis as características dos devedores e as suas condições econômicas e dos mercados vigentes, conforme a Figura 2.1 abaixo.

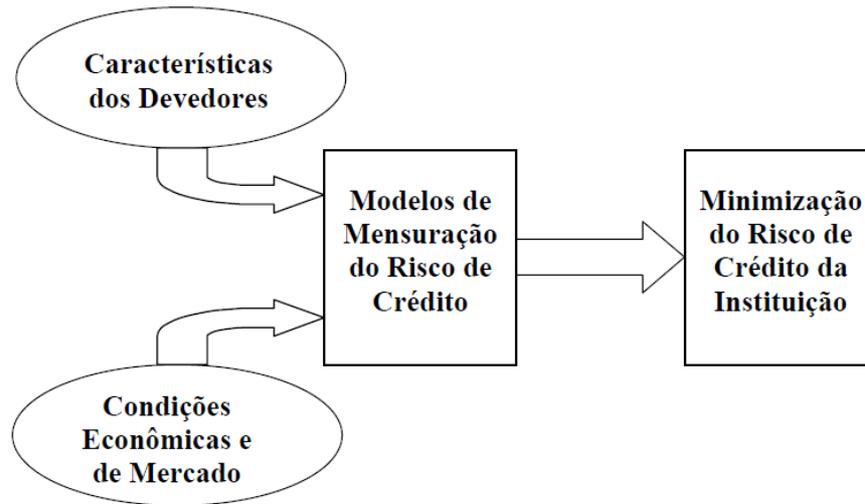


Figura 2.1 – Esquema de modelo de mensuração de risco

Características dos Devedores representam os parâmetros que instituições financeiras assumem para estabelecer a classificação dos clientes. Estas características normalmente são definidas pelas áreas de concessão de crédito que utilizam modelos de avaliação da capacidade creditícia dos clientes.

Condições Econômicas e de Mercado representam os parâmetros externos aos clientes que afetam a percepção de instituições bancárias sobre as probabilidades de inadimplência, servindo de base para apuração do risco de crédito. Estes parâmetros estão associados a variáveis macroeconômicas e de mercado ou dados históricos de inadimplência.

Eventos de inadimplência podem gerar perdas substanciais para as instituições financeiras. No entanto, modelos que permitem prever com antecedência este evento são utilizados na decisão de crédito e no monitoramento da evolução da qualidade de crédito.

Segundo Bessis (1998) o risco de crédito possui duas dimensões: a quantidade de risco e a qualidade do risco.

- A quantidade refere-se ao montante que pode ser potencialmente perdido nas operações de crédito;
- A qualidade refere-se a qual probabilidade ou aos sinais de possíveis perdas.

Diversas metodologias são utilizadas para avaliar o risco de crédito e classificar empresas em categorias de risco de acordo com a qualidade de crédito: modelos julgamentais, modelos empíricos que utilizam técnicas estatísticas e uma combinação dos dois. Essas metodologias vêm se aprimorando com o passar dos anos, especialmente com implementação de técnicas estatísticas mais desenvolvidas.

2.2 CREDIT SCORING

Este é um modelo de análise de concessão de crédito que tem como idéia básica identificar, através de análises estatísticas, os principais fatores que determinam a probabilidade de inadimplência. O modelo estabelece regras de pontuação (score) através de um conjunto de combinações de fatores que normalmente são as informações cadastrais dos clientes.

A metodologia básica para o desenvolvimento de um modelo de *credit scoring* não difere entre as aplicações para pessoa física ou jurídica. Como já mencionado, o objetivo do modelo é quantificar o risco associado ao crédito concedido, sendo, portanto uma ferramenta que auxilia na tomada de decisão de fornecer ou não crédito ao solicitante. Para construir o modelo de *credit scoring* existem alguns passos a serem seguidos, como mostra a Figura 2.2.

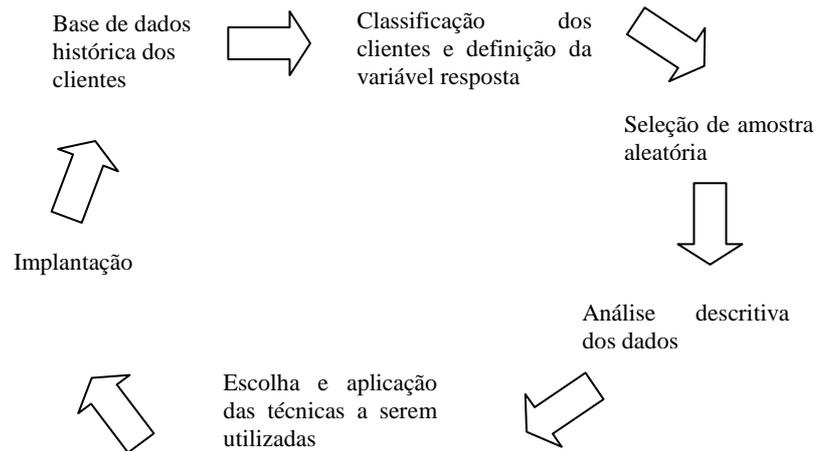


Figura 2.2 - Ciclo de desenvolvimento de um modelo

I. Levantamento de uma base histórica de clientes

A suposição básica para se construir um modelo de avaliação de crédito é que os clientes têm o mesmo padrão de comportamento ao longo do tempo; portanto, com base nas informações passadas são construídos os modelos. A disponibilidade e a qualidade da base de dados são fundamentais para o sucesso do modelo (TREVISANI *et al*, 2004).

II. Classificação dos clientes de acordo com o padrão de comportamento e definição da variável resposta

Nesta etapa são definidos quais são os clientes bons e quais são os clientes considerados maus pela instituição. Todavia, cabe observar que cada instituição tem sua própria política de crédito e estes conceitos de bons e maus pode mudar dependendo da instituição.

III. Seleção de amostra aleatória representativa da base histórica

Com a base de dados e a variável resposta definidas, são selecionadas amostras representativas de *bons clientes* e *maus clientes*. É importante que as amostras dos bons e maus clientes tenham o mesmo tamanho para se evitar um possível viés, devido à diferença de tamanhos.

IV. Análise descritiva e preparação dos dados

Consiste em analisar segundo critérios estatísticos cada variável a ser utilizada no modelo. Este tópico será abordado mais detalhadamente posteriormente.

V. Escolha e aplicação das técnicas a serem utilizadas para a construção do modelo

Existem diversas técnicas a serem utilizadas para a construção de modelos, algumas com maior ou menor complexidade. Neste trabalho serão utilizadas Regressão Logística

juntamente com a Lógica Fuzzy, na qual o procedimento metodológico será descrito no capítulo 4 deste trabalho.

VI. Implantação

Deve-se programar a implantação dos modelos. A instituição deve adequar seus sistemas para receber o algoritmo final e programar a utilização junto às demais áreas envolvidas.

Apesar de o *credit scoring* ser um processo matemático, ele não elimina a possibilidade de recusar um bom pagador e aceitar um mal pagador. Isto ocorre porque nenhum sistema de avaliação consegue capturar todas as informações relevantes que são necessárias para a precisa classificação dos devedores nas categorias e, mesmo que algum sistema conseguisse, o seu custo tornaria esse processo de análise economicamente inviável.

No capítulo a seguir apresenta-se os métodos de classificação Regressão Logística e Lógica *Fuzzy*, para a construção da metodologia para análise de concessão de crédito.

Capítulo 3

MÉTODOS

3.1 ANÁLISE DE REGRESSÃO

Uma das buscas da ciência é entender a associação entre variáveis. Isso porque entender tais associações pode ser útil de diversas maneiras, como na predição, ou seja, o conhecimento da associação entre variáveis pode fazer com que o comportamento de uma ou mais variáveis possa ser predito a partir do comportamento das variáveis relacionadas. Ainda é possível com tal conhecimento controlar o valor de uma variável a partir do ajuste das variáveis relacionadas.

De acordo com Royston & Sauerbrei (2008), modelos de regressão realizam muitas tarefas em todas as áreas da ciência onde dados empíricos são analisados, sendo que essas tarefas incluem:

- a) Predição de uma saída de interesse;
- b) Identificação de importantes preditores;
- c) Entendimento dos efeitos de preditores;
- d) Ajuste para preditores incontroláveis através de design experimental;
- e) Estratificação por risco.

Para Graybill & Iyer (2006), quase todas as decisões que um indivíduo toma são baseadas em predição e muitas dessas predições podem ser feitas através do estudo sistemático de associações e a análise de regressão trata do estudo dessas relações. Ainda segundo esses autores, há pelo menos duas razões pelas quais a predição é útil:

1. O valor verdadeiro da variável dependente Y é muito caro ou difícil para ser obtido, porém as variáveis predictoras são mais baratas ou fáceis de serem medidas;
2. A variável resposta é impossível de ser medida, frequentemente, por se tratar de valores futuros.

Para que uma predição neste sentido seja realizada, são necessários:

1. As variáveis preditoras, denotadas por X_1, X_2, \dots, X_p e os valores observados para essas variáveis;
2. Uma equação ou fórmula, para prever a variável resposta Y usando as variáveis preditoras, X_1, X_2, \dots, X_p .

De acordo com Ryan (2009), a Análise de Regressão é uma das técnicas estatísticas mais utilizadas, e seu uso está presente em quase todos os campos de aplicação. Weisberg (2005) afirma que a Análise de Regressão é a parte central de muitos projetos de pesquisa. Já para Johnson & Wichern (1992), análise de regressão é uma metodologia estatística para prever valores de uma ou mais variável resposta (dependente) a partir de uma coleção de valores de variáveis preditoras (independentes) e que também pode ser utilizada para avaliar os efeitos das variáveis preditoras nas respostas. Graybill & Iyer (2006), concordam com a definição de que a análise de regressão é um método comumente utilizado para obter uma função de predição para prever valores de uma variável resposta usando as variáveis preditoras.

O termo regressão foi proposto por Francis Galton em 1885, em um estudo onde demonstrou que a altura dos filhos não tende a refletir a altura dos pais, mas tende a regredir para a média da população (MAROCO, 2003 apud PREARO, 2008; FIGUEIRA, 2006). Para Johnson & Wichern (1992), o termo que foi escolhido a partir desse trabalho não reflete a importância nem a amplitude da aplicação desta metodologia.

3.1.1 Regressão Logística

A regressão logística é uma ferramenta estatística que tem sido bastante utilizada no desenvolvimento de modelos em *credit scoring*. A Regressão logística é uma técnica de análise multivariada, apropriada para as situações nas quais a variável dependente é categórica e assume um dentre dois resultados possíveis (binária), tais como “cliente bom ou cliente mau”. As variáveis independentes tanto podem ser categóricas quanto métricas.

O objetivo da regressão logística é gerar uma função matemática, cuja resposta permita estabelecer a probabilidade de uma observação pertencer a um grupo previamente determinado, em razão de um conjunto de variáveis independentes. Os coeficientes estimados pelo modelo de regressão indicam a importância da variável independente para a ocorrência do evento.

A regressão logística prediz a probabilidade de um evento ocorrer, a qual pode estar entre 0 e 1. A relação entre as variáveis independentes e dependentes se assemelha a uma curva em forma de S conforme ilustra o gráfico 1, a seguir.

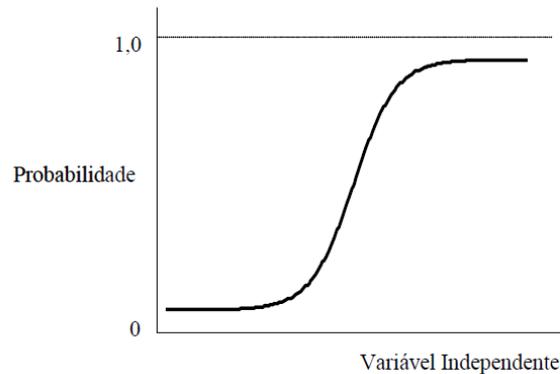


Figura 3.1 – Curva logística

Segundo Brito & Neto (2005), um aspecto que favorece a utilização da regressão logística é que seus resultados podem ser interpretados em termos de probabilidade. Esse fator se mostra particularmente importante nos modelos de risco de crédito, pois possibilita que seja medida a probabilidade de um determinado tomador assumir a condição de bom ou mau cliente, face em um conjunto de atributos.

3.1.1.2 Histórico

Segundo Lima (2002), a função logística surgiu em 1845, ligada a problemas de crescimento demográfico, problemas em que, até os dias de hoje, essa função é utilizada. Na década de 30, esta metodologia passou a ser aplicada no âmbito da biologia, e posteriormente nas áreas relacionadas a problemas econômicos e sociais. Paula (2002) aponta que, apesar do modelo de regressão logística ser conhecido desde os anos 50, foi devido a trabalhos do estatístico David Cox, na década de 70, que esta técnica se tornou bastante popular entre os usuários da Estatística.

A regressão logística vem sendo uma das principais ferramentas estatísticas utilizada na modelagem de diversos tipos de problemas. Paula (2002, p.118) explica:

Mesmo quando a resposta não é originalmente binária, alguns pesquisadores têm dicotomizado a variável resposta de modo que a probabilidade de sucesso possa ser modelada por intermédio da regressão logística. Tudo isso se teve, principalmente, à facilidade de interpretação dos parâmetros de um modelo logístico.

Segundo Hosmer e Lemeshow (2000, p.1):

Os métodos de análise de regressão tornaram-se um componente essencial de qualquer análise de dados que se preocupa em descrever a relação de uma variável dependente e uma ou mais variáveis explicativas. Frequentemente, a variável dependente discreta, assumindo dois ou mais possíveis valores. Na última década, o modelo de regressão logística, tornou-se, em muitos campos, o método padrão de análise nessa situação⁵.

Com base em vários estudos sobre mensuração de riscos, Dias Filho (2003, p.203) observou os principais fatores que guiaram a sua opção pela regressão logística:

- a) Comparada com outras técnicas de dependência, a regressão logística acolhe com mais facilidade variáveis categóricas. Aliás, esta é uma das razões pelas quais ela se torna uma boa alternativa à análise discriminante, sobretudo quando o pesquisador se defronta com problemas relacionados a variância;
- b) Mostra-se mais adequada a soluções de problemas que envolvem estimação de probabilidades, pois trabalha com uma escala que vai de zero a um;
- c) Requer um menor número de suposições iniciais, se comparada com outras técnicas utilizadas para discriminar grupos;
- d) Admite variáveis independentes métricas e não-métricas simultaneamente;
- e) Facilita a construção de modelos destinados à previsão de riscos em diversas áreas do conhecimento. Os chamados *Credit Scoring* e tantos outros;
- f) Tendo em vista que o referido modelo é mais flexível quanto às suposições iniciais, tende a ser mais útil e apresentar resultados mais confiáveis;
- g) Os resultados da análise podem ser interpretados com relativa facilidade, já que na lógica do modelo se assemelha em muito à de outras técnicas bem conhecida como a regressão linear.

3.1.1.3 Descrição do Modelo

Segundo Cordeiro (1986), a regressão logística consiste em relacionar, através de um modelo, uma variável resposta com fatores que influenciam na sua ocorrência. Mesmo quando a resposta de interesse não tem natureza binária, no caso deste trabalho, será estipulada uma probabilidade de sucesso para dicotomizar esta variável resposta (score de corte).

⁵ “Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. It is often the case that the outcome variable is discrete, taking on two or more possible values. Over the last decade the logistic regression model has become, in many fields, the standard method of analysis in this situation.”

Considerando a existência de apenas uma variável independente X_i (variável explicativa), temos como modelo a *Regressão Logística Simples*, sendo que sua forma usual é dada por:

$$E(Y_i | X_i) = \pi(X_i) = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}}, \quad (3.1)$$

Onde,

Y_i é a variável resposta binária p/ o i-ésimo indivíduo;

X_i é a variável independente p/ o i-ésimo indivíduo

$\beta = (\beta_0, \beta_1)$ é o vetor de coeficientes do modelo.

O modelo de regressão logística multivariado é uma extensão do modelo logístico simples, pois, a única diferença é que ao invés de utilizar só uma variável preditora X_i (independente ou explicativa), agora será utilizado duas ou mais variáveis explicativas X_1, X_2, \dots, X_{p-1} , sendo que o modelo é composto não só por variáveis preditoras, mas também dos coeficientes de regressão $\beta_0, \beta_1, \dots, \beta_{p-1}$ com a seguinte forma vetorial:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}_{p \times 1} \quad \mathbf{X} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_{p-1} \end{bmatrix}_{p \times 1} \quad \mathbf{X}_i = \begin{bmatrix} 1 \\ X_{i,1} \\ X_{i,2} \\ \vdots \\ X_{i,p-1} \end{bmatrix}_{p \times 1} \quad (3.2)$$

Logo temos:

$$\beta' \mathbf{X} = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} \quad (3.3.a)$$

$$\beta' \mathbf{X}_i = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_{p-1} X_{i,p-1} \quad (3.3.b)$$

Notação da função de resposta logística:

$$E(Y_i | X_i) = \pi(X_i) = \frac{\exp(\beta' \mathbf{X}_i)}{1 + \exp(\beta' \mathbf{X}_i)} \quad (3.4)$$

Com esta notação, a função de resposta logística simples estende-se à função de resposta logística múltipla. A variável dependente Y_i é dada por:

$$Y_i = E(Y_i | X_i) + \varepsilon_i \quad \text{ou} \quad Y_i = \pi(X_i) + \varepsilon_i \quad (3.5)$$

O termo ε_i é o erro do modelo e representa a diferença entre o valor observado de Y_i e o valor esperado condicionado de Y_i dado X_i , sendo Y_i uma variável dicotômica, assumindo apenas valores 0 ou 1, logo:

- Se $Y_i = 1 \Rightarrow \varepsilon_i = 1 - \pi(X_i)$, com probabilidade igual a $\pi(X_i)$.
- Se $Y_i = 0 \Rightarrow \varepsilon_i = -\pi(X_i)$, com probabilidade igual a $1 - \pi(X_i)$.

Assim ε_i tem distribuição com média “zero”, e variância $\pi(X_i)[1 - \pi(X_i)]$. Ou seja, a distribuição condicional da variável resposta segue uma distribuição binomial com probabilidade dada pela esperança condicional $\pi(X_i)$.

3.1.1.3.1 Estimação de Parâmetros

Os parâmetros desconhecidos do modelo a serem estimados são $\beta_0, \beta_1, \dots, \beta_{p-1}$. Esses parâmetros são estimados pelo método de Máxima Verossimilhança. Então é formulada a função de verossimilhança, que expressa as probabilidades das respostas observadas (que são independentes) em função dos parâmetros desconhecidos, dada por $l(\beta')$.

3.1.1.3.2 Função de Verossimilhança

Cada observação da variável aleatória Y_i tem uma distribuição de Bernoulli, onde:

$$P(Y_i = 1) = \pi(x_i) \text{ e } P(Y_i = 0) = 1 - \pi(x_i) \quad (3.6)$$

Então podemos representar sua distribuição de probabilidade, da seguinte forma:

$$f(Y_i) = \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i} \quad Y_i = 0 \text{ ou } 1 \quad i = 1, \dots, n \quad (3.7)$$

Logo, a função de verossimilhança será a seguinte:

$$l(\beta') = \prod_{i=1}^n f(Y_i) = \prod_{i=1}^n \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i} \quad (3.8)$$

Aplicando o logaritmo neperiano na função de verossimilhança e fazendo algumas operações numéricas, obtêm-se:

$$L(\beta') = \ln[l(\beta')] = \sum_{i=1}^n Y_i \beta' X_i - \sum_{i=1}^n \ln(1 + \exp(\beta' X_i)) \quad (3.9)$$

Para encontrar as estimativas de $\beta_0, \beta_1, \dots, \beta_{p-1}$ que maximizam a função (3.9), devem-se utilizar métodos numéricos, sendo que suas estimativas serão denotadas por $\hat{\beta}' = (b_0, b_1, \dots, b_{p-1})$. E a função resposta logística ajustada é dada por:

$$\hat{\pi}(X_i) = \frac{\exp(\hat{\beta}' X_i)}{1 + \exp(\hat{\beta}' X_i)} \quad (3.10)$$

3.1.1.3.3 Método de Escolha das Variáveis Explicativas do Modelo

Existem vários métodos de seleção das variáveis independentes relevantes para o modelo, o método utilizado para escolha de variáveis adotado no presente trabalho é denominado *forward stepwise*.

O método *forward stepwise* possibilita examinar um grande número de variáveis e, simultaneamente, examinar diversas equações de regressão logística possíveis, a partir dessas variáveis (VASCONCELOS 2002). Este método se inicia com um modelo sem nenhuma variável explicativa e a cada passo são incluídas as variáveis relevantes, até a obtenção do modelo final.

Para obtenção do modelo de regressão logística, a significância de cada variável é analisada com base no teste da razão de verossimilhança qui-quadrado (χ^2), assim, em qualquer passo do procedimento, a variável com maior poder de explicação é aquela que produz a maior variação na função de verossimilhança $L(\beta')$.

O software estatístico SPSS tem a opção “regressão logística, método *forward stepwise*”, que foi utilizada na determinação do modelo final de “*credit scoring*”. Os passos do procedimento *forward stepwise* estão resumidos a seguir:

Passo 0: Suponha a existência de “ p ” variáveis explicativas, o passo inicial computa um modelo contendo somente a constante (β_0) e sua função de log-verossimilhança L_0 . Em seguida, são computadas “ p ” modelos univariados contendo a constante mais uma variável explicativa (p modelos para p variáveis explicativas). Para cada um dos modelos contendo a variável X_1 , é computada a função de log-verossimilhança L_1^0 e o teste de razão de verossimilhança é determinado por uma estatística $G_1^0 = 2(L_1^0 - L_0)$, com distribuição χ^2 com $k-1$ grau de liberdade ($k =$ número de categorias da variável em questão). A variável mais relevante é aquela cujo p -value do teste é menor, e é denominada X_1^{0*} . O fato de uma

variável explicativa ter o menor p-value, não garante que essa variável é estatisticamente significativa, só será significativa se o p-value do teste for menor que o P_i (P_i é o nível de significância de inclusão da variável). Em seguida processo prossegue para o Passo 1.

Passo 1: Esse passo começa computando um modelo contendo a constante e a variável inclusa no passo anterior, e sua função de log-verossimilhança L_i^0 . Agora existem $p-1$ possíveis variáveis explicativas. Então são computados $p-1$ modelos contendo, cada um, a constante, a variável X_i^{0*} e mais uma variável explicativa ($p-1$ modelos para $p-1$ variáveis explicativas), e suas respectivas funções de log-verossimilhança. São realizados os testes de razão de verossimilhança para os $p-1$ modelos, $G_i^1 = 2(L_i^1 - L_i^{0*})$. A variável mais importante do passo 1 é aquela cujo p-value do teste é o menor, e é denominada X_i^{1*} . Se esse menor p-value for inferior a P_i , então o procedimento segue para o passo 2 contendo a constante, a variável X_i^{0*} , e a variável X_i^{1*} . Se o menor p-value for superior a P_i , então termina com o modelo contendo apenas constante e X_i^{0*} .

Passo 2: Essa etapa começa computando um modelo contendo a constante, X_i^{0*} e X_i^{1*} , e sua função de função de log-verossimilhança L_i^1 . Este passo inclui o teste para eliminação de variáveis, computando-se modelos e excluindo-se variáveis adicionadas em passos anteriores. Assim, são calculadas as funções de verossimilhança de um modelo removendo X_i^{0*} e de um modelo removendo X_i^{1*} , denominadas L_{-j}^{2*} . São realizados os testes da razão de verossimilhança, comparando-se os modelos sem cada uma das variáveis X_i^{0*} e X_i^{1*} ao modelo completo (contendo as duas variáveis e a constante), ou seja, $G_{-j}^2 = 2(L_i^1 - L_{-j}^{2*})$. A variável que deve ser eliminada nesse passo é aquela que fornecer o maior p-value. Esse maior p-value deve ser comparado ao nível de significância de exclusão (P_e) e, se for maior que P_e , então essa variável deve ser excluída. Após decidir sobre exclusão (ou não) da variável no Passo 2, o procedimento continua, no mesmo passo, com o processo de inclusão de variáveis, computando os modelos e adicionando-se cada uma das variáveis explicativas ainda não adicionadas e realizando os mesmos testes da razão de verossimilhança. Se nenhuma variável puder ser incluída, então o procedimento termina. Caso contrário, prossegue-se com a variável incluída no passo 0 (X_i^{0*}), no passo 1 (X_i^{1*}), sem a excluída no passo 2 e com a variável incluída no passo 2 (X_i^{2*}). Esse passo é repetido até se verificar a significância de todas as variáveis explicativas

Passo T: Esse passo terminal ocorre quando : a) todas as p variáveis estiverem no modelo ou b) todas as variáveis presentes no modelo tiverem p-value para exclusão menor que P_e e as variáveis não presentes no modelo tiverem p-value para inclusão maior que P_i .

Conforme Hosmer e Lemeshow (2000), muitos estudos sobre o procedimento stepwise mostram que adotar $P_i = 0,05$ (valor padrão em estudos estatísticos) é um valor decisório muito rigoroso, podendo deixar fora do modelo variáveis explicativas importantes. É recomendado que seja adotado P_i de 0,15 a 0,20, que garantiria a presença de variáveis importantes e garantia a seleção de variáveis com coeficientes significativamente diferentes de zero. O mesmo vale para a probabilidade de exclusão, P_e . No entanto, o valor escolhido de P_e deve ser maior que o valor de P_i , para evitar a possibilidade de incluir uma determinada variável em certo passo, e eliminá-la no passo subsequente. Tendo em vista essas observações, o modelo de credit scoring elaborado utilizou $P_i = 0,15$ e $P_e = 0,20$.

3.1.1.4 Razão de Chance

Uma das principais estatísticas utilizadas na análise de dados binários é a “razão de chances” (Agresti, 2003), onde a “chance” é definida como:

$$\frac{P(Y_i = 1 | X_i)}{P(Y_i = 0 | X_i)} = \frac{\pi(X_i)}{1 - \pi(X_i)}$$

A razão $\frac{\pi(X_i)}{1 - \pi(X_i)}$ com a transformação logit, é denominada como função resposta logit, e, p' é denominada de resposta logit. Considere a transformação logit da probabilidade:

$$\hat{p}_i' = \ln\left(\frac{\pi(X_i)}{1 - \pi(X_i)}\right) \quad (3.11)$$

Utilizando a (3.1), obtem-se a função ajustada para X_i :

$$\hat{p}_i' = b_0 + b_1 X_i \quad (3.12)$$

Considere o modelo da função ajustada (3.12) para $X_j = X_i + 1$, ou seja, acrescentando uma unidade no nível da preditora X_i , tem-se:

$$\begin{aligned} \hat{p}_j' &= b_0 + b_1 X_j \\ \hat{p}_j' &= b_0 + b_1 (X_i + 1) \end{aligned} \quad (3.13)$$

A diferença entre os dois valores ajustados será:

$$\hat{p}'_j - \hat{p}'_i = b_0 + b_1(X_i + 1) - (b_0 + b_1X_i) = b_1 \quad (3.14)$$

De acordo com (3.12), o \hat{p}'_i é o logaritmo neperiano da razão de chance estimada para X_i , denominada $\ln(\text{chance1})$. Do mesmo modo, \hat{p}'_j é logaritmo neperiano da chance estimada para $X_i + 1$, é denominado $\ln(\text{chance2})$. Assim, a diferença entre os dois valores ajustados pode ser dada por:

$$\ln(\text{chance2}) - \ln(\text{chance1}) = \ln\left(\frac{\text{chance2}}{\text{chance1}}\right) = b_1 \quad (3.15)$$

Tomando o exponencial de ambos os lados, tem-se:

$$\text{OR} = \text{Razão de Chances} = \left(\frac{\text{chance2}}{\text{chance1}}\right) = \exp(b_1) \quad (3.16)$$

Portanto, a razão de chances, $\exp(b_1)$, representa o aumento na probabilidade de $Y = 1$ quando o nível da preditora aumenta em uma unidade.

3.1.1.5 Diagnostico do Modelo Logístico

Para decidir qual modelo de regressão logística (credit scoring) será utilizado, é preciso aplicar alguns testes de validação desse modelo, é necessário verificar se há pontos influentes (a presença de outliers), se a função resposta é monotônica e em forma de S (sigmoidal), e se o modelo logístico ajustado é adequado.

3.1.1.6 Teste Wald de Bondade de Ajuste

Após a estimação dos parâmetros deve-se investigar a significância estatística dos mesmos no modelo. O teste de Wald é utilizado para avaliar se o parâmetro é significativo estatisticamente, a estatística deste teste é obtida pela razão do coeficiente pelo seu respectivo erro padrão:

$$W = \frac{\hat{\beta}}{EP(\hat{\beta})}, \quad (3.17)$$

sendo que W possui distribuição normal, e seu valor é comparado com valores tabulados, de acordo com o nível de significância definido, e sua hipótese nula é que os β 's são iguais a zero.

O Teste de Wald, frequentemente é falho, ou seja, rejeita coeficientes que são estatisticamente significativos para o modelo (HAUCK & DONNER, 1977). Assim aconselha-se que os coeficiente rejeitados pelo teste de Wald, sejam testados novamente pelo teste da razão de verossimilhanças. O teste da razão de verossimilhanças é obtido através da comparação entre os modelos com e sem as variáveis a serem testadas.

Suponha que existam $p+1$ parâmetros no modelo e que a hipótese nula estabelece que três sejam iguais a zero, a estatística (G), definida abaixo, compara o valor obtido com o de uma distribuição qui-quadrado com número de graus de liberdade igual à diferença no número dos parâmetros.

$$G = -2 \ln \left(\frac{\text{função de verossimilhança do modelo sem as variáveis}}{\text{função de verossimilhança do modelo com as variáveis}} \right). \quad (3.18)$$

3.1.1.6.1 Função de Deviance de Bondade de Ajuste

A Deviance (D) é uma estatística de bondade de ajuste muito enfatizada na literatura e baseia-se nas funções de log-verossimilhança maximizadas, sendo expressa por

$$D(y; \mu) = -2[l(\mu; y) - l(y; y)], \quad (3.19)$$

onde $l(y; y)$ representa a máxima verossimilhança de um ajuste exato, e $l(\mu; y)$ representa a máxima verossimilhança do modelo sob estudo.

Segundo Cordeiro (2004), a análise de deviance (ANODEV) é uma generalização da análise de variância para os modelos lineares generalizados, visando obter, a partir de uma seqüência de modelos, cada um incluindo mais termos que os anteriores. Dada uma seqüência de modelos encaixados, utiliza-se a deviance como medida de discrepância do modelo.

Sejam os modelos M_p e M_q ($p < q$) com “ p ” e “ q ” parâmetros, respectivamente. A estatística $D_p - D_q$ com “ $q - p$ ” graus de liberdade, é interpretada como uma medida de variação dos dados, explicada pelos termos que estão em M_q e não estão em M_p . Se $D_p - D_q > \chi^2(q - p; \alpha)$, considera-se que os termos que estão em M_q e não estão em M_p são significativos.

3.1.1.6.2 Gráfico do Delta Qui - Quadrado

O gráfico Delta Qui-quadrado versus a probabilidade estimada no j -ésimo nível do fator é usado para diagnosticar observações mal ajustadas (possíveis outliers).

O Delta Qui-quadrado é uma medida usada para calcular a mudança no qui-quadrado de Pearson provocada pela retirada de todas as observações com o j -ésimo nível de fator (covariável). Sua medida é usada para detectar níveis de fator mal ajustados. Esta medida é calculada como:

$$\chi_j^2 = \frac{\hat{r}_j^2}{1 - h_j} \quad (3.20)$$

onde:

\hat{r}_j^2 : resíduo de Pearson para o j -ésimo nível de fator;

h_j : j -ésimo elemento diagonal da matriz $H = X(X'X)^{-1}X'$, onde X é a matriz planejamento.

3.1.1.7 A Escolha do Score de Corte

Um fator que pode ser considerado na escolha do *score* de corte é puramente estatístico, baseado nas probabilidades de classificação incorreta de clientes. Nesse sentido, a idéia é minimizar as probabilidades de classificação incorreta, ou seja, minimizar a chance de classificar um cliente bom como ruim e a chance de classificar um cliente ruim como bom, simultaneamente. Com isso proposto, a fronteira de decisão que serve para auxiliar na definição do ponto de corte, ou seja, definir a partir de qual *score* será concedido o crédito para a linha de crédito em questão e estabelecida. Essa fronteira consiste em uma tabela que estima qual será o comportamento da carteira de crédito caso se adote determinado *score* de corte nas concessões de crédito. A fronteira é gerada com base nos *scores* estimados de todos os clientes contidos no banco de dados recebido.

3.2. LÓGICA *FUZZY*

A lógica *fuzzy* ou difusa foi estruturada em 1965 pelo Dr. Lofti A. Zadeh da Universidade da Califórnia para tratar e representar incertezas. A lógica difusa torna-se importante na medida em que o mundo em que vivemos não é constituído por fatos absolutamente verdadeiros ou falsos. É a lógica que permite representar valores de pertinência (grau de verdade) intermediários entre os valores de verdadeiro e falso da lógica clássica (bivalente). Pode ser aplicada, por exemplo, na construção de sistemas especialistas para descrever coisas imprecisas como: altura (alto, baixo), velocidade (rápido, lento), tamanho (grande, médio, pequeno), idade (jovem, velho).

3.2.1. Considerações Gerais

A lógica digital tradicional lida com variáveis assumindo apenas dois possíveis estados: falso ou verdadeiro. Em boa parte dos casos, esta representação é suficiente, mas há situações em que desejamos valores intermediários.

Poderíamos usar valores analógicos, mas aí cairíamos em equações matemáticas complexas que nem sempre nos implicariam no resultado desejado. Neste caso pode-se recorrer a lógica *fuzzy*. *Fuzzy*, em inglês, significa incerto, duvidoso. Expressa exatamente os valores com que lida. A lógica *fuzzy*, não trata uma variável como tendo apenas um estado atual, mas sim 'n' estados cada um com um grau de associação. Em outras palavras, não afirmamos que uma casa é grande, mas sim que ela é 0.8 grande, 0.2 média e 0.0 pequena. Isto faz com que definamos conjuntos em que um dado valor pode ser enquadrado. Voltando ao exemplo da casa, teríamos três conjuntos: casas grandes, médias e pequenas. Mas nada impede que tenhamos cinco conjuntos: casas enormes, grandes, médias, pequenas e minúsculas. O número de conjuntos nos diz quão precisamente estamos lidando com uma variável.

Assim como na lógica convencional, definimos regras nas quais associamos entradas para produzirem saídas. Podemos ainda utilizar os operadores OU e NÃO. Na lógica *fuzzy*, ao definirmos uma regra, informamos que quando uma variável assumir um dado conjunto E outra variável outro conjunto, teremos na saída tal variável com tal conjunto.

Finalmente, definimos um sistema *fuzzy*, que será uma coleção de variáveis de entrada, uma coleção de conjuntos para a variável de saída e uma coleção de regras que associam as entradas para resultar em conjuntos para a saída.

É necessária ainda uma função que "*desfuzzifique*" a saída, ou seja, que a partir dos graus de participação de cada variável de uma regra, retorne o grau de participação da saída e conseqüentemente o valor real da saída.

3.2.2. Descrição da Lógica Fuzzy

Seja X um espaço de objetos e x um elemento genérico de X . Um conjunto clássico A , $A \subseteq X$, é definido como uma coleção de elementos ou objetos $x \in X$, tal que cada x pode pertencer ou não ao conjunto A . Para definir uma função característica para cada elemento x em X , nós podemos representar um conjunto clássico A por um conjunto de pares ordenados $(x, 0)$ ou $(x, 1)$, os quais indicam $x \in A$ ou $x \notin A$, respectivamente (MORAIS, 2004).

Definição: Segundo MORAIS (2004), se X , é uma coleção de objetos, denotados genericamente por x , então, um conjunto *fuzzy* A em X , é um conjunto de pares ordenados, tal que:

$$A = (x, \mu_A(x)) / x \in X$$

Onde:

- A representa o conjunto *fuzzy*.
- x é o elemento pertencente ao universo de discurso X .
- X é o universo de discurso.
- $\mu_A(x)$ é a função de pertinência.

Os sistemas fuzzy são sistemas baseados em regras ou em conhecimento. A base consiste das chamadas regras fuzzy "Se-então", e está baseada em palavras e não em números, ou seja, quando o universo de discurso X é um espaço contínuo, nós usualmente dividimos o conjunto completo X em vários conjuntos *fuzzy*. Estes conjuntos, usualmente apresentam nomes que correspondem a adjetivos que aparecem em nosso uso lingüístico diário, e também são chamados de valores lingüísticos ou rótulos lingüísticos. Um conjunto *fuzzy* é unicamente especificado por sua função de pertinência, onde as regras e conclusões são obtidas em função deste valor, e não mais do valor propriamente dito da grandeza em questão (MORAIS, 2004).

A regra *fuzzy* "Se-então" é uma declaração na qual algumas palavras são representadas por uma função de pertinência (WANG, 1997, THE MATHWORKS, 1998, SHAW, SIMÕES, 1999, MIRANDA, 2000). Por exemplo:

- Proposição: **SE** a velocidade do carro é grande;
- Conclusão: **ENTÃO** aplicar uma força menor ao acelerador;

Onde **grande** e **menor** são representadas por funções de pertinência.

A lógica difusa (fuzzy) distingue-se da lógica booleana por permitir a utilização de um intervalo entre os limites 0 e 1, e não apenas estes, como no caso binário. Em outras palavras, podemos trabalhar não apenas com os valores 0 e 1, mas com todos os valores neste intervalo, ou seja, o intervalo [0 ; 1].

3.2.3. Funções de Pertinência Fuzzy mais Comuns

- Triangular

Este tipo de função serve para indicar apenas um único ponto de máximo no conjunto.

A definição é do tipo:

$$f(x) = \begin{cases} y_{\min} & x \leq i \quad \text{ou} \quad x \geq f \\ y_{\min} + (y_{\max} - y_{\min}) \cdot (x - i) / (m - i) & i < x \leq m \\ y_{\min} + (y_{\max} - y_{\min}) \cdot (f - x) / (f - m) & m < x < f \end{cases}$$

A Figura 3.2 mostra a função *fuzzy* triangular, onde o intervalo [i ; f] representa o lugar de existência da função.

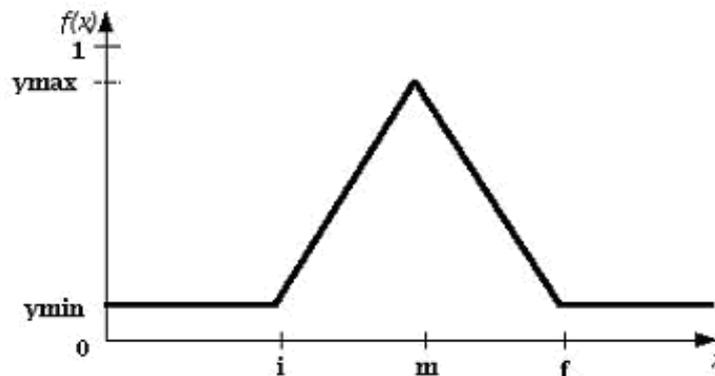


FIGURA 3.2 – FUNÇÃO FUZZY TRIANGULAR

Onde:

- y_{\min} é o menor valor possível para a função.
 - y_{\max} é o maior valor possível para a função.
 - i o início da ascendente da função.
 - m o valor de máximo.
 - f o final da descendente.
- Trapezoidal
- Neste tipo de função (Figura 3.3), é possível representar todo um intervalo de pontos de máximo. Podemos defini-la através da Equação 3.23.

$$f(x) = \begin{cases} y_{\min} & x \leq i \quad \text{ou} \quad x \geq f \\ y_{\min} + (y_{\max} - y_{\min}) \cdot (x - i) / (m1 - i) & i < x < m1 \\ y_{\max} & m1 \leq x \leq m2 \\ y_{\min} + (y_{\max} - y_{\min}) \cdot (f - x) / (f - m2) & m2 < x < f \end{cases} \quad \begin{array}{l} \text{Eq. 3.23 – Função} \\ \text{Trapezoidal} \end{array}$$

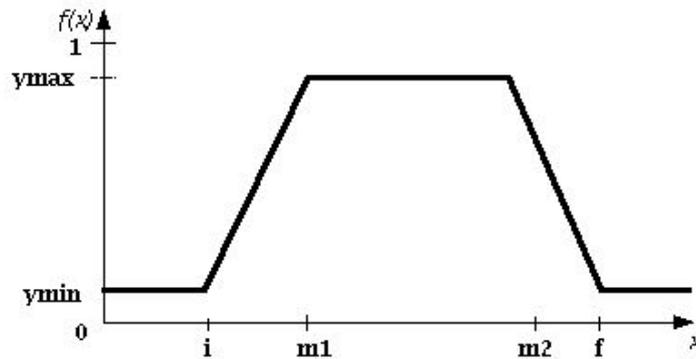


FIGURA 3.3 – FUNÇÃO FUZZY TRAPEZOIDAL

Onde:

- y_{\min} é o menor valor possível para a função.
 - y_{\max} é o maior valor possível para a função.
 - i o início do trapezio.
 - $m1$ o início do intervalo de máximo.
 - $m2$ o fim do intervalo de máximo.
 - f o final do trapézio.
- LR (Left-Right)

Este tipo de conjunto é mais genérico, pois permite a utilização de vários tipos de função para definir o intervalo de representação. As funções são chamadas de **L** (left, pois define a função à esquerda) e de **R** (right, pois define a função da direita), conforme mostra a figura 3.24

A formulação padrão é do tipo:

$$f(x) = \begin{cases} L((m1 - x) / \alpha) & x < m1 \\ 1 & m1 \leq x \leq m2 \\ R((m2 - x) / \beta) & x > m2 \end{cases} \quad \text{Eq. 3.24 – Função LR}$$

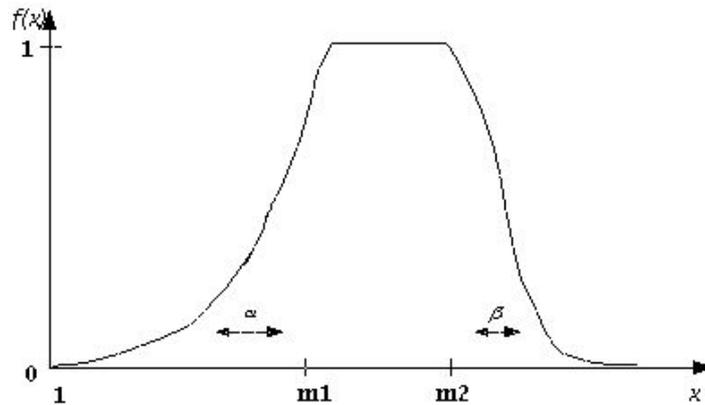


FIGURA 3.4 – FUNÇÃO FUZZY LR PARA $L = R = 1/(1 + X^2)$, ESPALHAMENTO ESQUERDO 5 E DIREITO 2.

Onde:

- L e R podem ser funções do tipo e-x, $1/(1+X^2)$, ou outra qualquer que tenha: $f(0) = f(1) = 0$ e mantenha-se no intervalo $[0;1]$.
 - $m1$ é onde inicia o intervalo de máximo.
 - $m2$ o fim do intervalo de máximo.
 - alfa é o espalhamento lateral da esquerda.
 - beta é o espalhamento lateral da direita.
- Gaussiana

As funções de pertinência Gaussianas são caracterizadas pela sua média (**m**) e seu desvio padrão (**K**). Este tipo de função de pertinência tem um decaimento suave e tem valores diferentes de zero para todo domínio da variável estudada (AMENDOLA et al., 2005). A Figura 3.5 exibe uma função de pertinência gaussiana. Nesta figura encontram-se no eixo vertical os valores da função de pertinência e no eixo horizontal os valores da variável que se quer estudar

$f(x) = e^{-k(x-m)^2} , k > 0$	Eq. 3.25 – Função Gaussiana
--------------------------------	------------------------------------

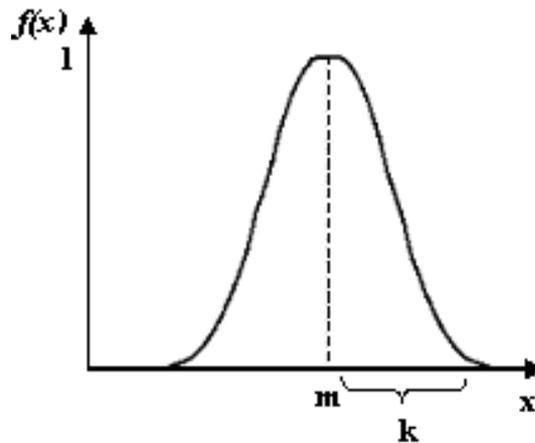


FIGURA 3.5 – FUNÇÃO FUZZY GAUSSIANA

Onde:

- K é o desvio padrão.
- m é a média da função.

3.2.4. Fuzzificadores e Defuzzificadores

Como na maioria das aplicações a entrada e a saída dos sistemas fuzzy são números reais, portanto devem-se implementar interfaces entre a máquina de inferência e o ambiente externo. Estas interfaces são o fuzzificador e o defuzzificador, (WANG, 1997, MATHWORKS, 1998, SHAW & SIMÕES, 1999, MIRANDA, 2000), conforme a Figura 3.6.

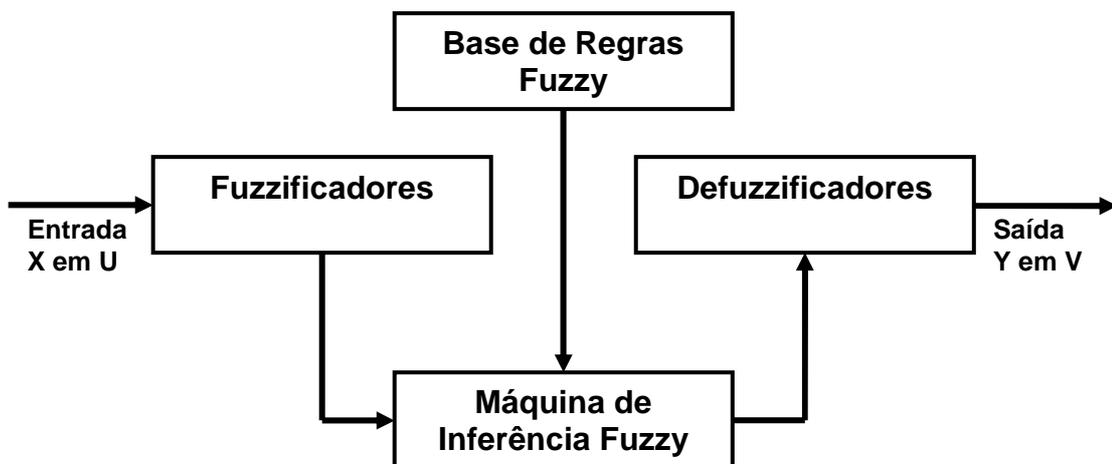


FIGURA 3.6 – ESTRUTURA DE SISTEMAS FUZZY

3.2.4.1 Fuzzificadores

Corresponde a etapa na qual os valores numéricos são transformados em graus de pertinência para um valor lingüístico. Cada valor de entrada terá um grau de pertinência em cada um dos conjuntos fuzzy. O tipo e a quantidade de funções de pertinência usados em um sistema dependem de alguns fatores tais como: precisão, estabilidade, facilidade de implementação, entre outras.

O fuzzificador é definido como um mapeamento de um ponto real $X^* \in U \subset R^n$ para um conjunto fuzzy A em U (WANG, 1997, MATHWORKS, 1998, SHAW & SIMÕES, 1999, MIRANDA, 2000).

3.2.4.1.1 Tipos de Fuzzificadores

- **Fuzzificador Singleton**

Mapeia um ponto real $X^* \in U \subset R$ em um fuzzy singleton A em U , que tem valor de pertinência 1 em X^* e 0 em todos os outros pontos em U , isto é (WANG, 1997, MATHWORKS, 1998, SHAW & SIMÕES, 1999, MIRANDA, 2000):

$\mu_A(X) = \begin{cases} 1 & \text{se } X = X^* \\ 0 & \text{se } X \neq X^* \end{cases}$	Eq. 3.26 – Fuzzificador Singleton
--	--

- **Fuzzificador Gaussiano**

Mapeia um ponto real $X^* \in U$ para um conjunto fuzzy A em U , que tem a seguinte função de pertinência:

$\mu_A(X) = e^{-\left(\frac{X_1 - X_1^*}{a_1}\right)^2} * \dots * e^{-\left(\frac{X_n - X_n^*}{a_n}\right)^2}$	Eq. 3.27 – Fuzzificador Gaussiano
--	--

Onde:

- a_i são parâmetros positivos.
- t -norma é usualmente escolhida como produto algébrico ou *min*.

- **Fuzzificador Triangular**

Mapeia um ponto real $X^* \in U$ para um conjunto fuzzy A em U , que tem a seguinte função de pertinência triangular:

$\mu_A(X) = \begin{cases} \left(1 - \frac{ X_1 - X_1^* }{b_1}\right) * \dots * \left(1 - \frac{ X_n - X_n^* }{b_n}\right) & \text{se } X_i - X_i^* \leq b_i; i = 1, 2, \dots, n \\ 0 & \text{outra condição} \end{cases}$	Eq. 3.28 Fuzzificador Triangular
---	---

Onde:

- a_i são parâmetros positivos.
- t -norma é usualmente escolhida como produto algébrico ou *min*

3.2.4.2. Defuzzificadores

Processo utilizado para converter o conjunto difuso de saída em um valor *crisp* correspondente.

O defuzzificador é definido como um mapeamento de um conjunto *fuzzy* B em $V \subset R$ para um ponto $Y^* \in V$ não-*fuzzy*. Conceitualmente, a tarefa do defuzzificador é especificar um ponto em V que melhor representa o conjunto fuzzy B (WANG, 1997, MATHWORKS, 1998, SHAW & SIMÕES, 1999, MIRANDA, 2000).

3.2.4.2.1 Tipos de Defuzzificadores

- **Defuzzificador Centro de Gravidade**

Especifica o Y^* como o centro da área coberta pela função de pertinência de B , com a seguinte equação:

$Y^* = \frac{\int_V Y \mu_B(Y) dY}{\int_V \mu_B(Y) dY}$	Eq. 3.29 – Defuzzificador Centro de Gravidade
---	--

- **Defuzzificador Centro Ponderado**

Como o conjunto fuzzy B é a união ou interseção de M conjuntos fuzzy, uma boa aproximação para o centro de gravidade é a ponderação dos centros dos M conjuntos fuzzy, com os pesos iguais às alturas dos correspondentes conjuntos fuzzy. Sendo Y^{-l} o centro do l ésimo conjunto fuzzy e W_l seu peso, o centro ponderado determina Y^* como:

$Y^* = \frac{\sum_{l=1}^M Y^{-l} W_l}{\sum_{l=1}^M W_l}$	Eq. 3.30 – Defuzzificador Centro Ponderado
--	---

- **Defuzzificador Máximo**

Este defuzzificador escolhe o Y^* como o ponto em V no qual centro da área coberta pela função de pertinência de B , ($\mu(B)$) atinge seu valor máximo. Definindo o conjunto:

$hgt(B) = \{ Y \in V / \mu_B(Y) = \sup_{Y \in B} \mu_B(Y) \}$	Eq. 3.31 – Defuzzificador Máximo
---	---

Se $hgt(B)$ contém mais de um ponto, pode-se usar uma das três opções:

- Menor dos Máximos: $Y^* = \inf \{ Y \in hgt(B) \}$
- Maior dos Máximos: $Y^* = \sup \{ Y \in hgt(B) \}$
- Média dos Máximos: $Y^* = \frac{\int_{hgt(B)} Y dY}{\int_{hgt(B)} dy}$

3.2.5. Conclusão

A lógica *Fuzzy* é uma importante ferramenta para auxílio na concepção de sistemas complexos, de difícil modelagem analítica, e pode ser utilizada em conjunto com outras técnicas não convencionais, como é o caso de uma combinação com as Redes Neurais Artificiais, os Algoritmos Genéticos, entre outras.

A implementação da lógica fuzzy vem se destacando nas diversas áreas, seja no reconhecimento de caracteres, em sistemas óticos, robôs, helicópteros comandados por voz através de um robô, sistemas de elevadores, aterrissagem de naves espaciais, entre outros.

Capítulo 4

DESCRIÇÃO DO ESTUDO

4.1 ASPECTOS GERAIS

Uma instituição financeira do ramo de cartões de crédito deseja conceder empréstimos a seus clientes e, para isso, necessita de uma ferramenta que avalie o grau de risco associado a cada empréstimo para auxiliar no processo de tomada de decisão. Para viabilizar o processo, foram disponibilizadas informações do histórico de clientes usuários do serviço.

4.2 POPULAÇÃO E AMOSTRA

Neste trabalho serão utilizados os dados cadastrais de clientes – usuários de cartão de crédito – de uma Instituição financeira, coletadas no período de 03/01/2000 a 29/09/2003, totalizando um universo de clientes. Deste universo de clientes retirou-se uma amostra de 5.917 clientes considerando estimativas com o nível de confiança de 95% e um erro amostral de 1%. A determinação do tamanho da amostra para populações finitas foi obtida aplicando a fórmula abaixo:

$$n = \frac{z^2 \cdot p \cdot q \cdot N}{e^2(N-1) + z^2 \cdot p \cdot q}$$

Onde:

n = tamanho da amostra;

N = tamanho da população;

Z = 1,96, valor correspondente à probabilidade central de 95% na tabela da distribuição normal padrão;

p = proporção de sucesso, neste caso a que a observação venha ser de um “*bom cliente*”.

q = proporção de fracasso ($1-p$), neste caso a que a observação venha ser de um “*mal cliente*”.

e = 0,01, erro amostral.

4.3 DEFINIÇÃO DE VARIÁVEIS

Para realização deste trabalho foram utilizadas 13 variáveis, sendo que 6 dessas variáveis são qualitativas e 7 variáveis são denominadas como quantitativas, elas são definidas a seguir:

- SALÁRIO – O salário do cliente;
- QT_DEP – A quantidade de dependentes que o cliente possui;
- VL_CONTRATO – O valor do contrato;
- QT_PARCELAS – A quantidade de parcelas;
- QT_PENDENCIAS – A quantidades de pendências financeiras (SERASA) do cliente;
- QT_CHEQ_DEV – A quantidade de cheques devolvidos emitido pelo cliente;
- GRAU_INST – O grau de instrução do cliente;
 - 0 – Analfabeto;
 - 1 – Fundamental Completo;
 - 2 – Ensino Médio Incompleto;
 - 3 – Ensino Médio Completo;
 - 4 – Superior Incompleto;
 - 5 – Superior Completo;
- RESIDÊNCIA – O cliente possui residência própria?
 - 0 – Não;
 - 1 – Sim;
- SEXO – O sexo do cliente;
 - 0 – Feminino;
 - 1 – Masculino;
- EST_CIV – O estado civil do cliente;
 - 0 – Solteiro;
 - 1 – Casado;
 - 2 – Divorciado;
 - 3 – Viúvo;
- REG_CAS – O regime de casamento do cliente;
 - 0 – Separação de bens;
 - 1 – Comunhão parcial de bens;
 - 2 – Comunhão total de bens;
 - 3 – Não informado;

- QUAL_CREDIT – A qualidade de crédito do cliente;
0 – Ruim;
1 – Boa;
- IDADE – A idade do cliente.

4.4 METODOLOGIA

Esta seção será apresentada os passos para construção dos modelos.

Passo 1: Preparação da base de dados

Para o desenvolvimento do modelo, foi definida, em primeiro lugar, a variável dependente “qualidade de crédito”, Criou-se uma variável categórica de 1 a 3, com valor 1 para clientes adimplentes (“bom”) e 0 para clientes inadimplentes (“ruim”). A seguir, definiram-se as variáveis explicativas disponíveis e relevantes para a construção do modelo, sendo que 7 delas são contínuas ou discretas e 5 são do tipo categorica, conforme apresentação das variáveis na seção anterior.

Passo 2: Regressão logística

Aplicou-se regressão logística, utilizando o software SPSS 17.0 (Statistical Package for the Social Sciences), considerando os níveis de significância para entrada e saída de variáveis iguais a 5% e 10%, respectivamente. No presente caso, após o método stepwise, o modelo final para discriminação dos três grupos de clientes deveria conter apenas as variáveis

Passo 3: Lógica Fuzzy

O modelo fuzzy foi desenvolvido utilizando o pacote Fuzzy Logical Toolbox 2.1 do programa Matlab® 7.8.0 (The MathWorks, Inc. – R2009a). O Matlab se caracteriza por um sistema interativo feito através de uma “Janela de Comando”, onde o usuário fornece os comandos e os resultados são apresentados. A linguagem de programação utilizada é o Fortran, largamente usada em ciências da computação e análise numérica. Pode-se fazer distintos usos do Matlab com seus vários aplicativos Toolbox, que são reuniões de arquivos destinados a tratar de diferentes problemas. Dentre eles, o Fuzzy Logic Toolbox disponibiliza arquivos e funções destinados a algumas aplicações da teoria de conjuntos fuzzy. O pacote Fuzzy Toolbox permite particularmente o desenvolvimento de modelos

lingüísticos baseados em lógica fuzzy, oferecendo duas alternativas de métodos de inferência, Sugeno e Mamdani.

Todas as metodologias propostas nesta seção serão discutidas em forma de resultados no capítulo a seguir.

Capítulo 5

APLICAÇÃO

5.1 Aspectos gerais

Neste capítulo, dividi-se em três momentos:

- *Análise exploratória dos dados;*

A análise exploratória dos dados nesta secção consiste apenas em uma ferramenta de análise descritiva para observar as relações entre as variáveis explicativas e a variável resposta do modelo.

- *Aplicação das Técnicas de classificação;*
 - a. Regressão Logística;
 - b. Análise discriminante;
 - c. Lógica Fuzzy;
- *Análise de desempenho das técnicas;*

5.2 Análise Exploratória dos dados

O conjunto de informações (banco de dados) foi dividido em dois grupos, grupo de treinamento (representa 70% dos dados – 4.142 cadastros de clientes) e o grupo de validação (representa 30% dos dados – 1.775 cadastros de clientes).

O banco em estudo é referente a um conjunto de dados de clientes – usuários de cartão de crédito – de uma Instituição financeira, extraída no período jan./2000 a set./2003, com 5.917 clientes cadastrados nesse período. A variável de interesse desse estudo é a “*qualidade de crédito*”, que classifica os clientes em “*bom*” ou “*ruim*”, para concessão de crédito, que dependerá de alguns atributos (as variáveis explicativas ou independentes).

Tabela 5.2.1 - Distribuição dos Clientes quanto a Qualidade de Crédito

	Clientes	%
Ruim	1.254	21,2%
Bom	4.663	78,8%
Total	5.917	100,0%

Observa-se na Tabela 5.2.1, que dos 4.142 clientes desta instituição, 78,8% são bons clientes, e os 21,2% restante são considerados clientes ruins.

Quanto ao gênero, a Tabela 5.2.2 mostra que os clientes do gênero masculino possuem uma grande e desfavorável incidência quanto à qualidade de crédito, pois é maioria neste grupo de maus clientes (73,4%), porém, analisando somente a categoria de clientes bons, verifica-se que os clientes do gênero feminino é maioria neste grupo (56,7%).

Tabela 5.2.2 - Distribuição do Gênero do Cliente quanto à Qualidade de Crédito

	Qualidade de Crédito		Total
	Ruim	Bom	
Masculino	921 73.4%	2.019 43.3%	2.940 49.7%
Feminino	333 26.6%	2.644 56.7%	2.977 50.3%
Total	1.254 100,0%	4.663 100,0%	5.917 100,0%

A Figura 5.2.1 revela a condição da qualidade de crédito, quanto ao estado civil do cliente. Observou-se que dentre os “*maus clientes*”, observou-se que 45,8% são casados, 43,9% são solteiros e 10,3% são divorciados. Dentre os “*bons clientes*”, observou-se que 43,0% são casados, 46,6% são solteiros, 10,1% são divorciados e apenas 0,3% são viúvos. Percebe-se, também, que entre os dois grupos não há muita divergência do percentual de ocorrência de cliente “*bom*” ou “*ruim*” entre os solteiros, divorciados e viúvos.

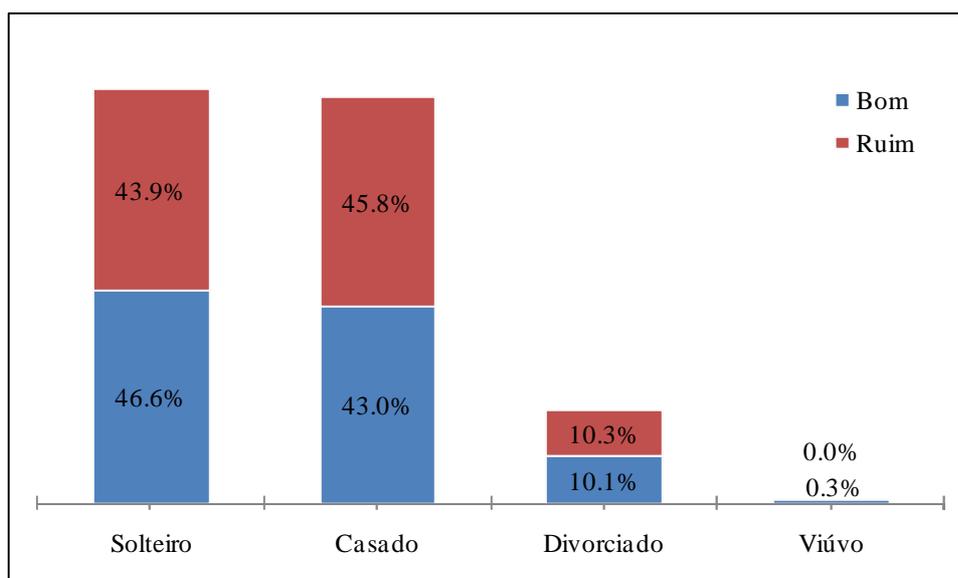


Figura 5.2.1 – Distribuição do Estado Civil do Cliente quanto a Qualidade de Crédito

Observa-se na Figura 5.2.2, que quanto aos clientes ruins, os maiores percentuais encontram-se entre os clientes com o Ensino Fundamental (37,9%), seguido dos que são do

grupo dos Analfabetos (29,5%). Comparando os grupos quanto a qualidade de crédito, verifica-se que os clientes com o Ensino Médio Incompleto destacam-se com a maior diferença de percentuais, ou seja, 22,8% são clientes bons contra 12,8% de clientes ruins (diferença igual a 10,0 pontos percentuais), seguido dos Analfabetos com a diferença de percentual, apresentando 29,5% de clientes ruins contra 21,3% Clientes bons (diferença igual a 8,2 pontos percentuais).

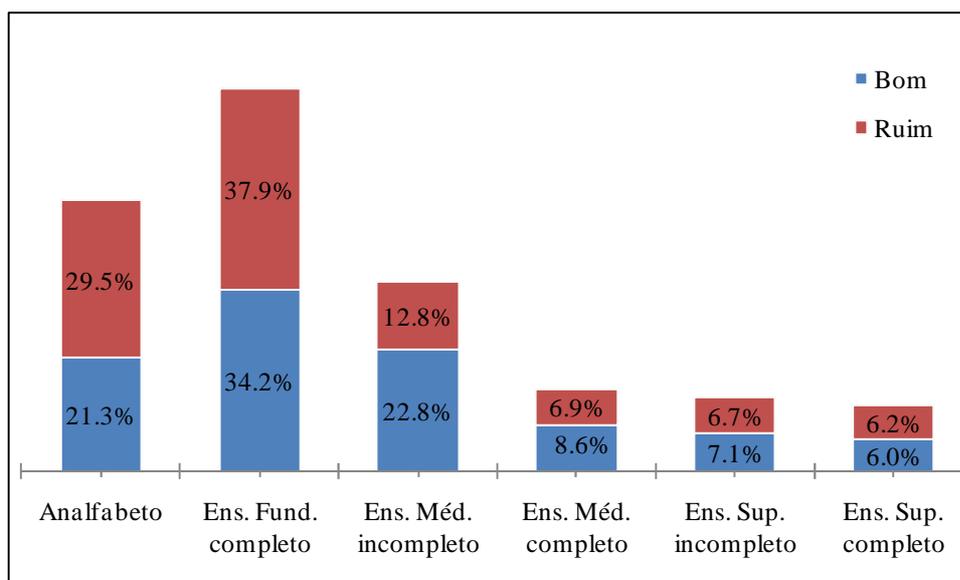


Figura 5.2.2 – Distribuição do Grau de Escolaridade do Cliente quanto a Qualidade de Crédito.

Observando o cruzamento da variável regime de casamento quanto a variável qualidade de crédito, a Tabela 5.2.3 apresenta que dentre os clientes ruins, o maior percentual encontra-se entre os clientes com regime de casamento por separação de bens com 64,8% e somente 35,2% por comunhão total de bens, o mesmo acontecendo dentre os clientes bons, com 68,0% dos clientes possuem regime de casamento por separação de bens e 32,0% por comunhão total de bens.

Tabela 5.2.3 - Distribuição do regime de casamento do cliente quanto à Qualidade de Crédito

	Qualidade de Crédito		Total
	Ruim	Bom	
Separção de Bens	813 64,8%	3.171 68,0%	3.982 67,3%
Comunhão total de bens	441 35,2%	1.492 32,0%	1.935 32,7%
	1.254	4.663	5.917
Total	100,00%	100,00%	100,00%

Observa-se na Figura 5.2.3, que dentre os clientes ruins, o maior percentual encontra-se entre os clientes com residência própria 90,4% e somente 9,6% dos clientes com residência

não própria, o mesmo acontecendo dentre os clientes bons, 93,0% dos clientes possuem residência própria e 7,0% com residência não própria.

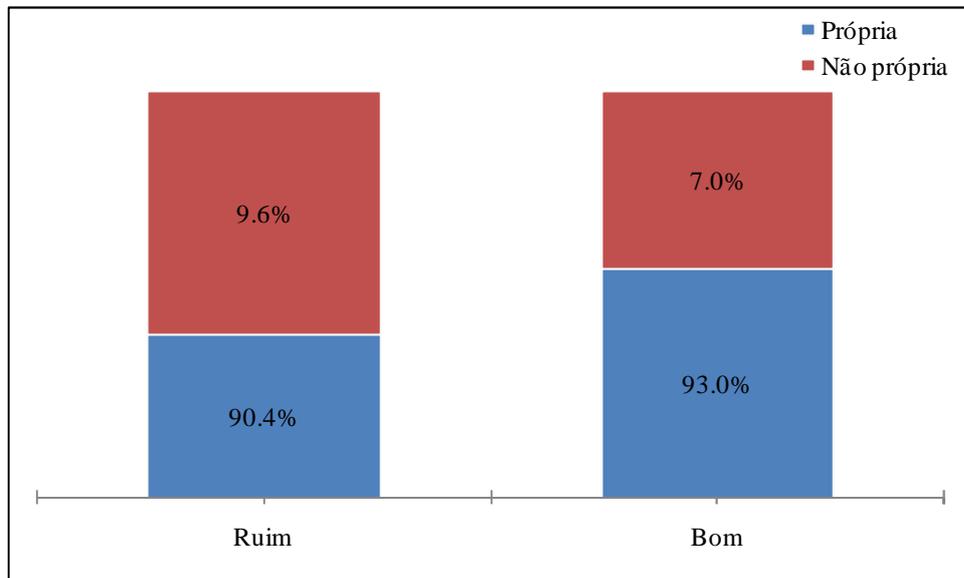


Figura 5.2.3 – Distribuição da situação de residência do cliente quanto à Qualidade de Crédito

Quanto à restrição do cliente ao Serasa, a Figura 5.2.4 apresenta que dentre os clientes que possuem restrição junto ao Serasa 48,0% deste são considerados “*Cientes Ruins*”, e dentre os clientes que não possuem Restrição junto ao Serasa 80,2% é composta pela classe de “*Cientes Bons*”, pode-se observar também que há uma semelhança (não grande diferença) da proporção de clientes que possuem restrição junto ao Serasa nas duas classes quanto à qualidade de crédito.

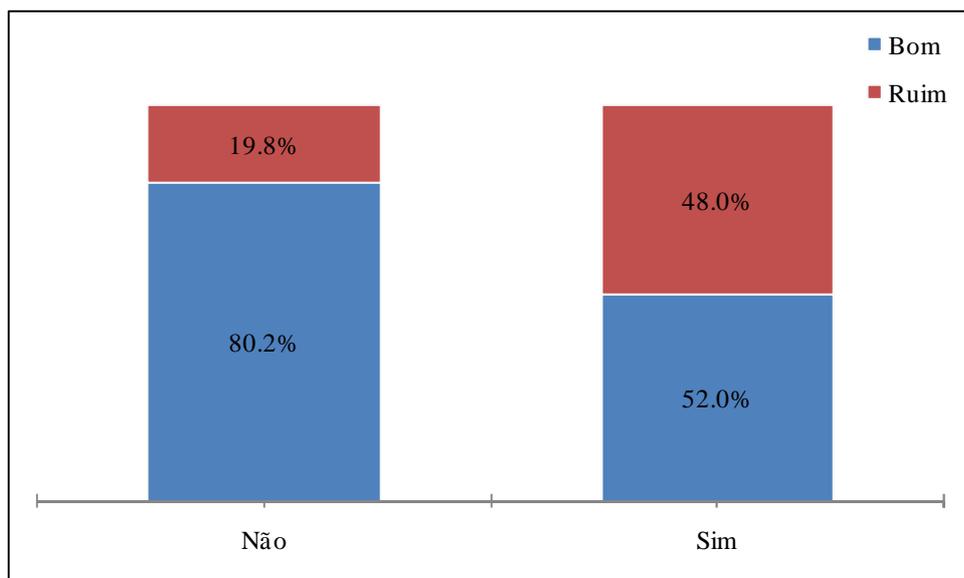


Figura 5.2.4 - Distribuição da Restrição do cliente no Serasa quanto à Qualidade de Crédito.

A Tabela 5.2.4 mostra que, a média salarial dos Clientes bons é de R\$ 1.720,71 e dos clientes ruins é R\$ 1.492,79, ou seja, a análise amostral sugere que em média os clientes bons

têm renda superior àqueles que são clientes ruins, os valores mínimos e máximos da renda mensal dos clientes bons correspondem a R\$ 203,00 e R\$ 7.995,00, respectivamente, enquanto que os salários dos clientes ruins variam de R\$ 203,00 a R\$ 7.946,00. Observa-se que 25% dos clientes bons possuem salários menores que R\$ 1.214,00 e 25% dos clientes bons possuem salários maiores que R\$ 2.066,00. Analisando os clientes ruins, temos que 25% dos clientes ruins possuem salários menores que R\$ 981,00 e 25% possuem salários maiores que R\$ 1.809,00.

Tabela 5.2.4 – Resumo Estatístico do Salário Mensal (em R\$), quanto à qualidade crédito dos Clientes.

	Qualidade de Crédito						Desvio Padrão
	Média	1º Quartil	Mediana	3º Quartil	Mínimo	Máximo	
Ruim	1.492,79	981,00	1.343,00	1.809,00	203,00	7.946,00	920,45
Bom	1.720,71	1.214,00	1.567,00	2.066,00	203,00	7.995,00	893,61

Observando-se a Tabela 5.2.5, entre os clientes com “*boa conduta*”, têm-se que em média possuem idade de 40 anos, o cliente com menor idade tem 19 anos, e com idade máxima, 78 anos. Enquanto os clientes com “*má conduta*”, em média possuem a idade de 39 anos, cujas observações mínimas e máximas correspondem a 20 e 80 anos, respectivamente, e também, observa-se que 75% dos “*Bons clientes*” possuem a idade abaixo de 47 anos e 75% dos “*Maus clientes*” possuem a idade inferior a 44 anos.

Tabela 5.2.5 – Resumo Estatístico da Idade, quanto à qualidade crédito dos Clientes.

	Qualidade de Crédito						Desvio Padrão
	Média	1º Quartil	Mediana	3º Quartil	Mínimo	Máximo	
Ruim	39	32	37	44	20	80	10
Bom	40	33	39	47	19	78	11

A Tabela 5.2.6 apresenta que o valor de contrato médio dos clientes bons é de R\$ 2.482,85, com valor mínimo de R\$ 240,00 e máximo de R\$ 11.920,00, já o contrato médio dos clientes ruins é de R\$ 2.795,85, com valor mínimo e máximo de R\$ 240,00 e R\$ 12.000,00, respectivamente. Observa-se também, que 25% dos clientes bons possuem o valor do contrato abaixo de R\$ 1.090,00 e 25% dos clientes bons possuem o valor de contrato superior a R\$ 3.290,00. Analisando os clientes ruins, temos que 25% dos clientes ruins possuem o valor de contrato menor que R\$ 1.407,00 e 25% dos clientes possuem o valor do contrato maior que R\$ 3.600,00.

Tabela 5.2.6 – Resumo Estatístico do Valor do Contrato (em R\$), quanto à qualidade crédito dos Clientes.

	Qualidade de Crédito						Desvio Padrão
	Média	1º Quartil	Mediana	3º Quartil	Mínimo	Máximo	
Ruim	2.795,85	1.407,00	2.328,50	3.600,00	240,00	12.000,00	1.936,28
Bom	2.482,85	1.090,00	2.000,00	3.290,00	240,00	11.920,00	1.871,64

A Tabela 5.2.7 mostra que, a quantidade média de parcelas dos clientes bons é de 15 parcelas, já para os clientes ruins é de 21 parcelas, cujas observações mínimas e máximas, tanto para clientes bons e ruins, correspondem a 1 e 24 parcelas respectivamente, observa-se também, entre os clientes de má qualidade de crédito, uma quantidade média de parcelas alta em relação aos clientes com boa qualidade. Pode-se dizer que, 25% dos clientes bons possuem uma quantidade de parcelas inferior a 10 e 25% possuem uma quantidade de parcelas igual a 24. Já os clientes de má qualidade, 75% destes possuem uma quantidade de parcelas superior a 18.

Tabela 5.2.7 – Resumo Estatístico do Número de Parcelas, quanto à qualidade crédito dos Clientes.

	Qualidade de Crédito						Desvio Padrão
	Média	1º Quartil	Mediana	3º Quartil	Mínimo	Máximo	
Ruim	21	18	24	24	1	24	6
Bom	15	10	12	24	1	24	8

5.3. Regressão Logística

O modelo logístico de *Credit Scoring* será utilizado nesta seção para estimar, a partir do conhecimento de uma série de variáveis, a probabilidade de um cliente ser classificado no grupo de *bons credores*, utilizando-se como variável resposta a *qualidade de crédito*, conforme discutido no capítulo anterior.

Esses resultados compreendem a seleção das variáveis explicativas significativas no modelo final, coeficientes estimados dessas variáveis e os testes de hipótese realizados para validação deste modelo, bem como o estudo de um Score de corte ótimo.

5.3.1. Construção do Modelo Logístico

A base de dados de treinamento (corresponde a 70% das informações, pois a base foi dividida, sendo 30% desses dados será utilizado para validação do modelo) utilizada para o ajuste do modelo logístico contém o registro de 4.142 clientes, dos quais 77,9% são classificados como “*clientes bons*” e 22,1% como “*clientes Ruins*”. Originalmente o banco possui 13 variáveis, sendo 01 variável dependente e 12 preditoras (05 variáveis qualitativas e

07 quantitativas), representando as informações pessoais e financeiras dos clientes. Para as variáveis qualitativas, foram criadas variáveis indicadoras, ou seja, cada variável preditora terá $(p-1)$ variáveis indicadoras – p é o número de fatores de cada variável preditora – para que possibilite que o efeito no ajuste por conta de cada resposta do cliente.

Como um dos objetivos desta análise é identificar quais variáveis são mais eficientes na categorização dos dois tipos de clientes bancários, um procedimento *stepwise* foi empregado. O método de seleção escolhido foi o *forward stepwise* (Hosmer e Lemeshow, 1989). Resumidamente, o procedimento se inicia através da estimação de um modelo apenas com o intercepto, inclui, uma a uma, as variáveis mais significantes no modelo e exclui aquelas que, na presença das outras, não são mais importantes. O procedimento termina quando nenhuma variável puder ser incluída ou excluída no modelo de acordo com níveis de significância preestabelecidos.

O modelo foi estimado através do software *SPSS 17.0 WINDOWS*, sendo adotados $P_i = 0,05$ e $P_e = 0,10$, valores críticos da probabilidade de inclusão e da probabilidade de exclusão a cada passo do método de seleção das variáveis. O software SPSS utiliza o teste escore para escolher a variável a ser incluída. Para a exclusão, foi usado o teste de razão de verossimilhança (LR – *likelihood-ratio test*). As variáveis indicadoras resultantes de uma mesma variável foram tratadas de forma independente. Assim, é possível a inclusão de apenas algumas das variáveis indicadoras de uma determinada variável. No desenvolvimento de um modelo para a utilização em uma instituição financeira, o procedimento *stepwise* é o primeiro passo para a obtenção do modelo final.

Substituições de variáveis e recategorizações são freqüentemente feitas para diminuir a multicolinearidade, tornar o modelo mais interpretável e melhorar o desempenho. Foi feita pequenas alterações na variável *Grau de Escolaridade*, ou seja, houve uma junção de fatores (como mostra a Tabela 5.3.1), para que houvesse significância em todas suas variáveis indicadoras (ou *variáveis dummies*).

Tabela 5.3.1 - Recategorização das variável “*Grau de instrução*”.

Grau de Instrução	
Original	Recategorizada
1 – Analfabeto	1 - Analfabeto
2 - E. Fund. completo	2 - E. Fund. completo
3 - E. Méd. incompleto	
4 - E. Méd. completo	3 - E. Méd. completo
5 - E. Sup. incompleto	
6 - E. Sup. completo	4 - E. Sup. completo

5.3.2. Estimação dos Parâmetros

Com objetivo de avaliar fatores que influenciam a qualidade de crédito, as variáveis selecionadas no modelo logístico, em seu formato categorizado (todas em formato de variável *dummy*), estão na Tabela 5.3.2, onde são mostradas as variáveis categorizadas e seus respectivos coeficientes estimados.

Tabela 5.3.2 – Variáveis e Coeficientes Estimados no Modelo Ajustado.

		B	S.E.	Wald	GL	p-valor.	OR (Razão de Chance)	(%) Vantagem
Grau de Instrução	<i>Analfabeto</i>	- 0,887	0,194	20,977	1	0,000	0,412	-58,8%
	<i>Ensino Fundamental</i>	- 0,484	0,183	6,970	1	0,008	0,617	-38,3%
	<i>Ensino Médio</i>	- 0,731	0,213	11,826	1	0,001	0,481	-51,9%
	<i>Ensino Superior*</i>	-	-	26,675	3	0,000	-	-
Residência própria	<i>Não*</i>	-	-	-	0	-	-	-
	<i>Sim</i>	- 0,482	0,160	9,110	1	0,003	0,617	-38,3%
Sexo do Cliente	<i>Masculino</i>	- 1,307	0,097	180,751	1	0,000	0,271	-72,9%
	<i>Feminino*</i>	-	-	-	0	-	-	-
Pendência no Serasa	<i>Não</i>	1,045	0,173	36,623	1	0,000	2,845	184,5%
	<i>Sim*</i>	-	-	-	0	-	-	-
Faixa de Idade	<i>18 ---/ 39*</i>	-	-	42,364	2	0,000	-	-
	<i>39 ---/ 60</i>	0,588	0,099	35,254	1	0,000	1,800	80,0%
	<i>60 ---/ 81</i>	0,880	0,227	15,082	1	0,000	2,412	141,2%
Faixa do número de dependente	<i>Não</i>	- 1,908	0,152	157,122	1	0,000	0,148	-85,2%
	<i>Sim*</i>	-	-	-	0	-	-	-
Faixa da Quantidade de Parcelas	<i>Até 7</i>	3,050	0,217	198,290	1	0,000	21,119	2011,9%
	<i>7 ---/ 13</i>	1,491	0,110	183,509	1	0,000	4,444	344,4%
	<i>13 ---/ 18</i>	1,557	0,219	50,465	1	0,000	4,744	374,4%
	<i>18 ---/ 24*</i>	-	-	337,260	3	0,000	-	-
Possui cheques devolvidos	<i>Sem cheques</i>	0,757	0,142	28,506	1	0,000	2,133	113,3%
	<i>Possui Cheques devolvidos</i>	-	-	-	0	-	-	-
Faixa de Salário mínimo	<i>Até 6*</i>	-	-	37,718	3	0,000	-	-
	<i>6 ---/ 11</i>	- 0,232	0,105	4,920	1	0,027	0,793	-20,7%
	<i>11 ---/ 16</i>	- 0,796	0,142	31,251	1	0,000	0,451	-54,9%
	<i>Acima de 16 SM</i>	- 0,777	0,201	14,886	1	0,000	0,460	-54,0%
	Constante	0,849	0,298	8,100	1	0,004	2,336	133,6%

* É a combinação linear das outras categorias (Dummy)

5.3.2.1. Teste de significância de cada variável

A estatística de Wald é definida como o quadrado da razão entre o coeficiente logístico estimado e o seu erro padrão. Por meio desta estatística, que tem distribuição Qui-quadrado, testa-se a seguinte hipótese estatística para cada variável independente:

H_{0i} : o coeficiente logístico é igual a zero.

H_{1i} : o coeficiente logístico é diferente de zero.

A Tabela 5.3.2 apresenta os coeficientes de regressão, as estatísticas de Wald, odds-ratio e respectivos intervalos de confiança para cada um dos 15 parâmetros. de acordo com o critério de Wald, todos os parâmetros submetem-se à exigência de um nível de significância de 5%, estimados pelo método de stepwise.

A Tabela 5.3.2 apresenta o conjunto de variáveis que compõem o modelo ajustado junto com suas respectivas categorias que foram significativas no modelo testadas pela estatística Wald (*teste de significância individual dos parâmetros*), e comprovado pelo p-valor dessa estatística. Assim, constata-se que os coeficientes de todas as variáveis incluídas no modelo logístico são estatisticamente diferentes de zero (rejeitando-se assim hipótese H_{0i} , p-valor < 0,05). Apresenta também as estatísticas Odds-Ration (Razão de chance) e % vantagem que cada fator (categoria da variável) sobrepõe em relação ao grupo de risco (*clientes ruins*).

5.3.1.2. Teste de significância do modelo – Teste de bondade de ajuste

Observa-se na Tabela 5.3.3, que as estatísticas de bondade de ajuste são unânimes na aceitação do modelo, pois os métodos *Deviance*, *Pearson* e *Hosmer-Lemeshow* indicam com p-valor > 0,05 que há evidências suficientes para aceitação do modelo.

Tabela 5.3.3 - Teste de Bondade de ajuste de Pearson, Deviance e Hosmer-Lemeshow.

Método	Qui-quadrado	Gráus de Liberdade	P-Valor
Hosmer - Lemeshow	13,101	8	0,324
Pearson	688,202	679	0,395
Deviance	695,260	679	1,000

5.2.1.3. Teste G:

A qualidade de ajuste do modelo escolhido pode ser verificada também pela estatística do Teste G, usada para testar as hipóteses:

H_0 : Todos os coeficientes do modelo são nulos;

H_1 : Pelo menos um coeficiente é não nulo;

Assim o resultado $G = 1.071,192$, e 16 graus de liberdade com um nível descritivo (P -valor < 0,05), indicando que existe evidência suficiente que no mínimo um dos coeficientes é diferente de zero.

5.2.1.4. Análise da Razão de chance:

Nesta seção é apresentada análise da razão de chance $\left(\frac{\text{Cliente Bom}}{\text{Cliente Ruim}}\right)$ para algumas das variáveis dispostas na Tabela 5.3.2.

- Um Cliente sem Pendência no Serasa, possui aproximadamente 2 vezes (a probabilidade de 184,5%) mais chance de ser classificado com boa qualidade de crédito, do que um cliente com má qualidade de crédito.
- Um Cliente com idade entre 39 e 60 anos, tem 80% de chance de ser classificado com boa qualidade de crédito, do que um cliente com má qualidade de crédito.
- Um Cliente analfabeto tem aproximadamente 60% de perder em chance de ser classificado com boa qualidade de crédito, do que um cliente com má qualidade de crédito.

5.2.1.5 Ponto de corte e Validação do modelo ajustado

Quando não se conhece a proporção populacional, costuma-se usar o ponto de corte 0,5 que define probabilidades iguais para os dois grupos. Um modelo adequado deve identificar a porcentagem mais alta dos casos classificados corretamente e assim encontra-se o ponto de corte (k) ótimo. Para o modelo ajustado pelo banco de treinamento, o ponto de corte igual $k = 0,50$ foi que obteve maior percentual de acertos de classificação, isto é, 82,6% dos dados foram classificados corretamente.

Para conseguir a eficiência classificatória do modelo, a amostra foi separada em duas partes: uma utilizada para estimação do modelo (amostra com 4.142 – 70% dos dados), e outra para testar a eficiência da classificação – *Amostra de validação* – 1.775 clientes (30% dos dados). A amostra utilizada para estimação, também chamada de amostra de treinamento.

A validação do modelo de Regressão logística é obtida através da aplicação do modelo na amostra de validação (Hair, 1998). As taxas de acertos na amostra de validação são quase idênticas às taxas de acerto da amostra de treinamento (como mostra a Tabelas 5.3.4 e a Tabela 5.3.5). Isto leva conclusão de que o modelo de regressão logística possui forte suporte empírico tanto na amostra de validação quanto na amostra de treinamento.

Tabela 5.3.4 - Taxa de acerto do modelo logístico da Amostra de Treinamento.

		Qualidade de Crédito (predito)		
		Ruim	Bom	% Correto
Qualidade de Crédito (observada)	Ruim	331	567	36,9%
	Bom	154	3.090	95,3%
Taxa de Acerto do modelo				82,6%

* Valor de corte é 0,50

Tabela 5.3.5 - Taxa de acerto do modelo logístico da Amostra de Validação.

		Qualidade de Crédito (predito)		
		Ruim	Bom	% Correto
Qualidade de Crédito (observada)	Ruim	110	246	30,9%
	Bom	79	1.340	94,4%
Taxa de Acerto do modelo				81,7%

* Valor de corte é 0,50

5.4. Lógica Fuzzy

Para a construção do modelo fuzzy foi utilizada a metodologia desenvolvida pelo modelo lingüístico fuzzy do tipo Mamdani, com suporte do pacote Fuzzy Logical Toolbox 2.1 do programa Matlab® 7.8.0 (The MathWorks, Inc. – R2009a). O modelo utiliza um método de inferência fuzzy associado a um conjunto de regras do tipo “*If-Then*”, com a característica de permitir a utilização do conhecimento prévio sobre as regras. O conseqüente (*então*) de cada regra é obtido pela interseção fuzzy (*t*-norma, sendo utilizado o mínimo) das variáveis de entrada e seu valor representa o grau de ativação da regra. A saída final do modelo é um número real, uma média ponderada das regras de acordo com o grau de ativação, com valores entre 0 e 1, grau de pertinência da situação conseqüente segundo as condições estabelecidas.

A Tabela 5.4.1 e a Figura 5.4.1, apresenta o resumo dos parâmetros do sistema no uso da ferramenta.

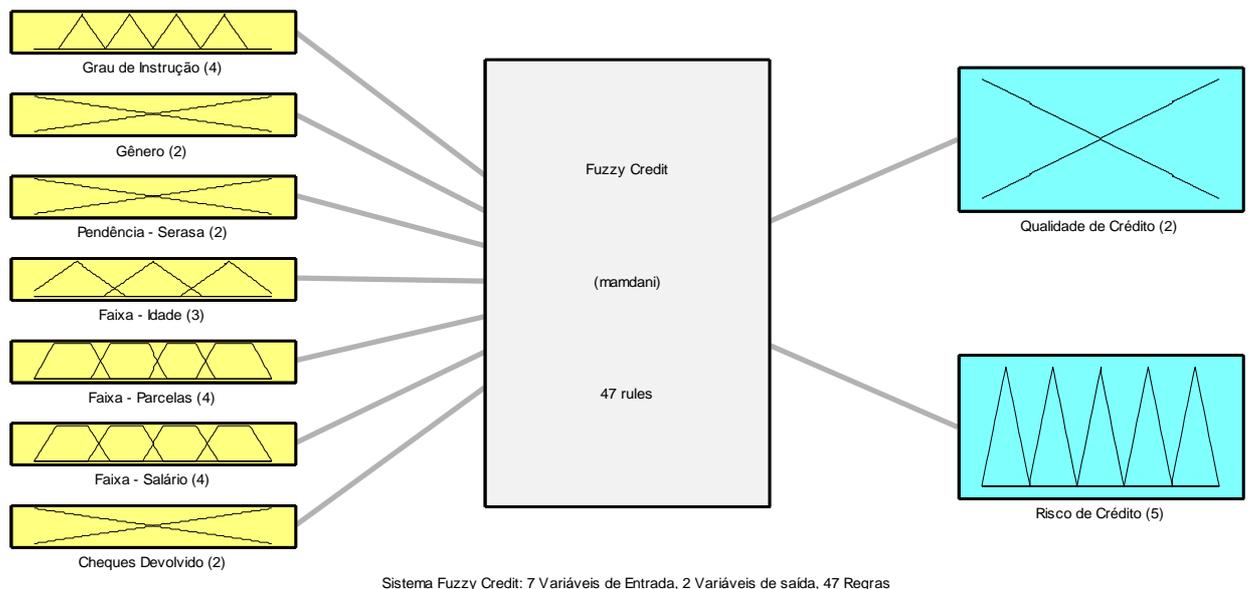


Figura 5.4.1 – Resumo da Ferramenta do Sistema Fuzzy

Tabela 5.4.1 - Resumo dos Parâmetros do Sistema fuzzy

Type	<i>mamdani'</i>	<i>Onde a resposta do processo é um conjunto difuso para cada regra</i>
andMethod	<i>min'</i>	<i>Utilizado para ser o conector das regras do sistema</i>
orMethod	<i>max'</i>	<i>Utilizado para ser o conector das regras do sistema</i>
defuzzMethod	<i>bisector'</i>	<i>Por ser mais aderente nos testes de simulação</i>
input:	<i>[1x7 struct]</i>	<i>7 variáveis de entrada</i>
output:	<i>[1x2 struct]</i>	<i>2 variável de saída</i>
Rule	<i>[1x47 struct]</i>	<i>47 regras no total</i>

Para melhor sensibilidade e adequação da modelagem fuzzy utilizou 7 variáveis de entrada (input) – variáveis significativas no modelo ajustado que melhor discriminam o cliente quanto a qualidade de crédito analisada pela razão de chance (ver Tabela 5.3.2) – e 2 variáveis de saída (output), o tratamento numérico das variáveis do modelo fuzzy com base no conhecimento adquirido com o método da regressão logística (ver seção 5.3).

Tabela 5.4.2 – Variáveis lingüísticas do Sistema Fuzzy

Variável Lingüística		Conjuntos Fuzzy
Input	Grau de Instrução	<i>Analfabeto</i>
		<i>Ensino Fundamental</i>
		<i>Ensino Médio</i>
		<i>Ensino Superior</i>
	Sexo do Cliente	<i>Masculino</i>
		<i>Feminino</i>
	Pendência no Serasa	<i>Não</i>
		<i>Sim</i>
	Faixa de Idade	<i>18 ---/ 39</i>
		<i>39 ---/ 60</i>
		<i>60 ---/ 81</i>
	Faixa da Quantidade de Parcelas	<i>Até 7</i>
		<i>7 ---/ 13</i>
		<i>13 ---/ 18</i>
<i>18 ---/ 24</i>		
Possui cheques devolvidos	<i>Sem cheques</i>	
	<i>Possui Cheques devolvidos</i>	
Faixa de Salário mínimo	<i>Até 6</i>	
	<i>6 ---/ 11</i>	
	<i>11 ---/ 16</i>	
	<i>Acima de 16 SM</i>	
Output	Qualidade de Crédito	<i>Ruim</i>
		<i>Bom</i>
	Risco de Crédito	<i>Muito Baixo</i>
		<i>Baixo</i>
		<i>Médio</i>
		<i>Alto</i>
	<i>Muito Alto</i>	

A partir das 7 variáveis de entrada é estabelecida duas variáveis de saída que expressam, o nível da *qualidade de crédito* – com termos lingüístico: *Ruim* e *Bom* – e o Risco associado ao crédito (*Risco de crédito*) – com termos lingüístico: *Muito baixo*, *Baixo*, *Médio*, *Alto* e *Muito alto* – alcançado pelo cliente (ver Figura 5.4.2 e Figura 5.4.3), a partir da combinação (perfil) das característica do cliente.

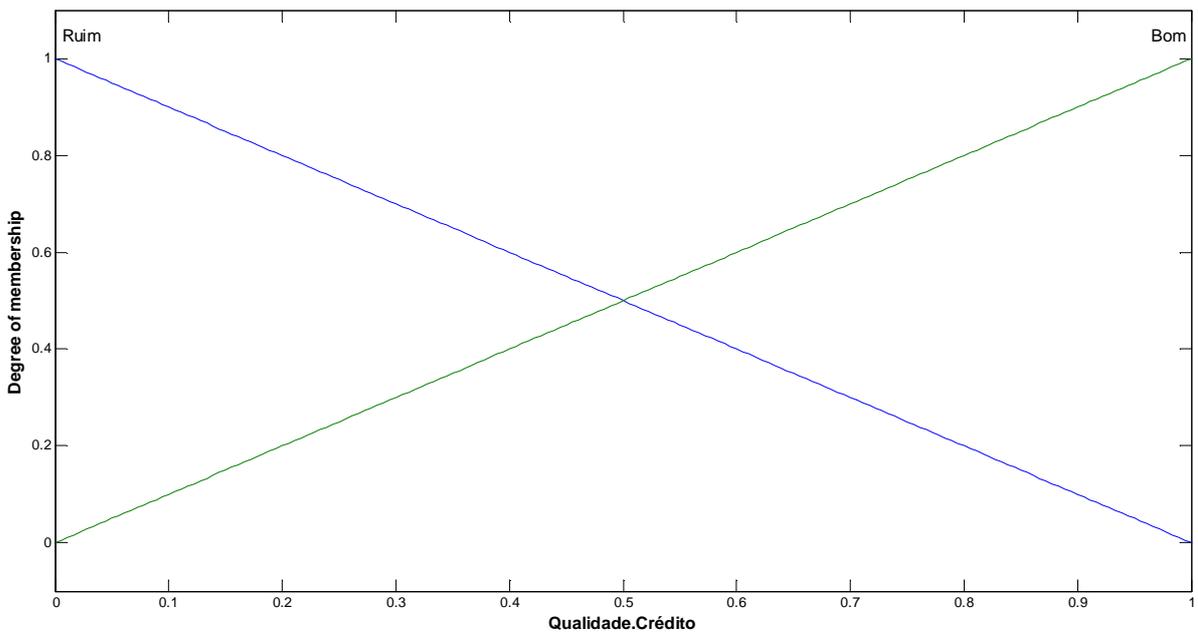


Figura 5.4.2 – Função de pertinência associada a variável de saída “*Qualidade de Crédito*”

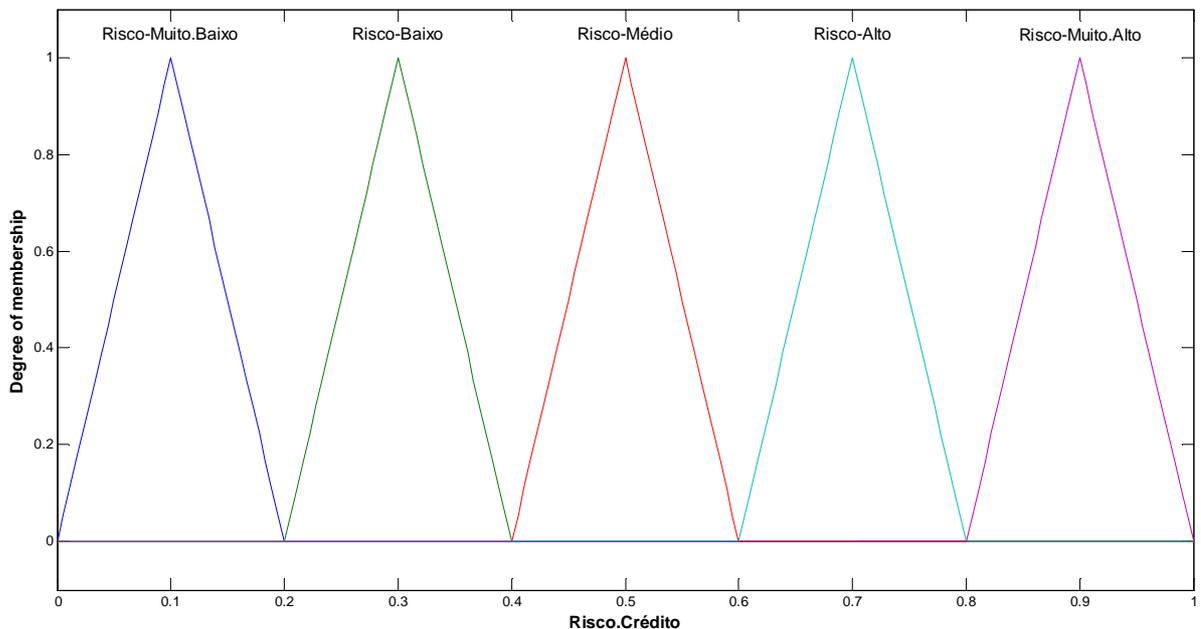


Figura 5.4.3 – Função de pertinência associada a variável de saída “*Risco de Crédito*”

A Figura 5.4.4 mostra a estrutura lingüística do funcionamento do modelo utilizado, para tal processo, onde ocorre a inferência, existem entradas escalares para cada variável que se transforma em um conjunto de graus de pertinências cujo vetor é empregado para limitar os

conjuntos fuzzy de saída das variáveis – *Qualidade e Risco de crédito*, baseada no conjunto de regras especificadas no sistema, apresentadas a seguir:

1. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1).
2. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 60---|81) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
3. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
4. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
5. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
6. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
7. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
8. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
9. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
10. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Muito.Alto) (1)
11. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
12. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
13. If (Grau.Inst is Ens.Med.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
14. If (Grau.Inst is Ens.Sup) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
15. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 13_---|18) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
16. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)

17. If (Grau.Inst is Ens.Sup) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
18. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
19. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is Mais.de.16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
20. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Muito.Baixo) (1)
21. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
22. If (Grau.Inst is Ens.Sup) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
23. If (Grau.Inst is Ens.Sup) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
24. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
25. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 13_---|18) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
26. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
27. If (Grau.Inst is Ens.Med.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
28. If (Grau.Inst is Ens.Med.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
29. If (Grau.Inst is Ens.Sup) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Baixo) (1)
30. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Médio) (1)
31. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Médio) (1)
32. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Médio) (1)
33. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Médio) (1)

34. If (Grau.Inst is Ens.Sup) and (Gênero is Feminino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Médio) (1)
35. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Não) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is Até-7) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Não.Possui) then (Qualidade.Crédito is Bom)(Risco.Crédito is Risco-Médio) (1)
36. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 7---|13) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
37. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
38. If (Grau.Inst is Analf.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is Até.6) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
39. If (Grau.Inst is Ens.Med.) and (Gênero is Feminino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
40. If (Grau.Inst is Ens.Med.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 18|---|39) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
41. If (Grau.Inst is Ens.Fund.) and (Gênero is Masculino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 6---|11) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)
42. 42. If (Grau.Inst is Ens.Fund.) and (Gênero is Feminino) and (Pendência.Serasa is Sim) and (Faixa.Idade is 39---|60) and (Faixa.Parcelas is 18_---|24) and (Faixa.Salário is 11---|16) and (Cheque.Dev is Possui) then (Qualidade.Crédito is Ruim)(Risco.Crédito is Risco-Alto) (1)

Sendo assim, foi selecionado aleatoriamente o perfil de um cliente da base de dados (utilizada para executar a modelagem logística), com o seguinte perfil: *Analfabeto (1)*, *Sexo Masculino(0)*, *60 ano de idade (2)*, *com renda mensal de 3 salários mínimos (1)*, *com restrição no Serasa (1)*, *com dívida parcelada em 24 vezes (4)* e *possui a emissão de cheques que foram devolvidos (1)*, pelo processo de inferência fuzzy, conclui-se que esse cliente tem um risco de 90% (Risco muito alto) de ser um inadimplente, por conseqüência a qualidade de seu crédito é considerada Ruim (levando em consideração o ponto de corte adota na modelagem logística – 0,50).

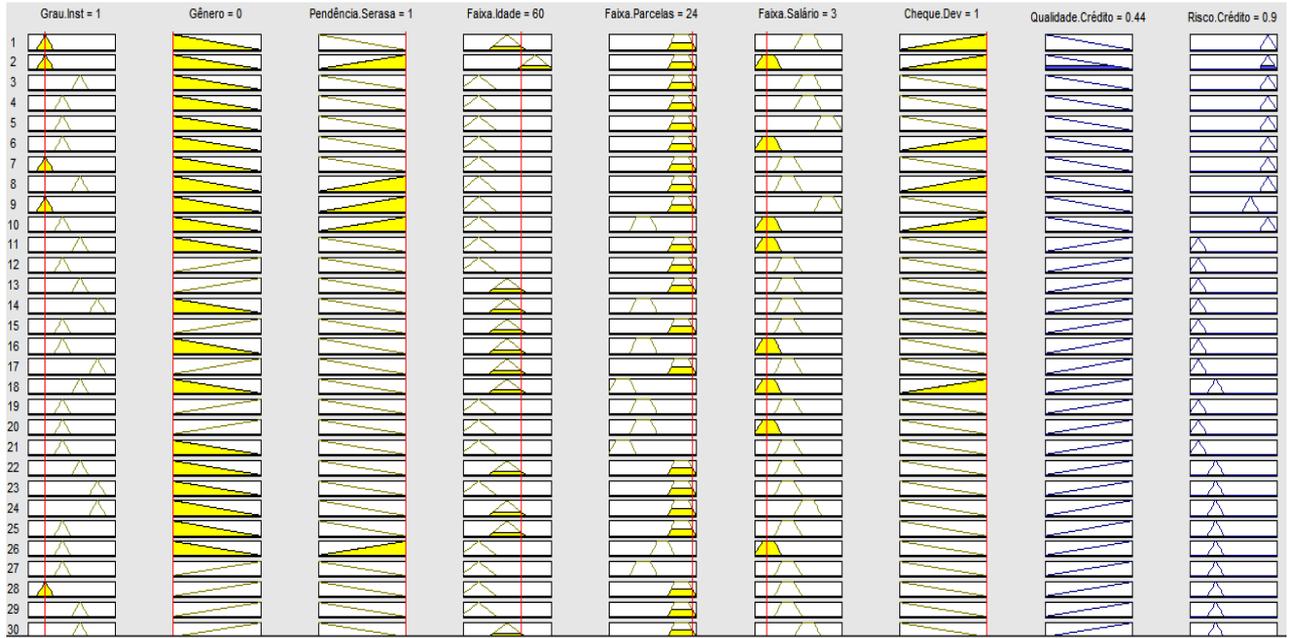


Figura 5.4.4 – Funcionamento do Modelo Fuzzy para Análise de Crédito.

A partir do conhecimento adquirido da modelagem da regressão logística, foi possível desenvolver um Modelo Fuzzy para análise de crédito, tendo como resposta a os Score da qualidade e o risco associado ao crédito.

Capítulo 6

Considerações Finais e Perspectivas

No presente estudo, foi desenvolvida uma metodologia para análise de concessão de crédito utilizando a técnica estatística de *Regressão Logística* integrada ao método de inteligência artificial *Lógica fuzzy*.

Em situações onde não se tenha a presença de um especialista pode-se utilizar a regressão logística para auxiliar na construção das regras a serem utilizadas na lógica Fuzzy.

Apesar da utilização da Lógica Fuzzy no Sistema de Tomada de Decisão para concessão de um bem de crédito, ter cumprido com êxito os objetivos propostos neste trabalho, é possível que a contribuição desta técnica seja ainda mais relevante se for utilizada em conjunto com outras técnicas de inteligência artificial, como Redes Neurais, Algoritmos Genéticos, ou Redes Bayesianas. Essas técnicas poderiam auxiliar na etapa de criação das regras utilizadas na Lógica Fuzzy, permitindo assim, a montagem de regras de forma dinâmica, as quais poderiam variar ao longo do tempo. Desta forma, teríamos sistemas híbridos, onde seria possível aproveitar ao máximo os benefícios de cada técnica. Pretende-se também expandir essa metodologia para outros segmentos que sente a necessidade de ter um “agente inteligente”, como na indústria de processos, na área de marketing, na saúde, etc...

Referências Bibliográficas

AGRESTI, A. **Categorical data Analysis**, Second Edition, John Wiley and Sons: New York. 2003.

AKIAMA, S. R. Probabilidade de inadimplência de grandes empresas no sistema financeiro nacional. Dissertação de Mestrado. Universidade de São Paulo. 2008.

ALTMAN, E.I. (1968). **Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy**. *Journal of Finance*, 23, 589-609.

ASSAF NETO, Alexandre, TIBÚRCIO SILVA, Cesar Augusto. **Administração de capital e giro**. 2ª Ed. São Paulo: Atlas, 1999.

BESSIS, J. **Risk management in banking**. New York: John Wiley & Sons, 1998.

BLATT, A. **Avaliação de risco e decisão de crédito: um enfoque prático**. São Paulo: Nobel, 1999.

Brito, Giovani A. S.; Neto, Alexandre A. **Modelo De Classificação De Risco De Crédito De Grandes Empresas**. Disponível: <<http://www.congressosp.fipecafi.org/artigos52005/383.pdf>>. Acesso em: 24 de Fev. 2010

CAOUILLE, JOHN B., ALTMAN, Edward. I. e NARAYANAN, Paul. **Managing Credit Risk – The next Great Financial Challenge**, New York: John Wiley & Son Inc., 1998

CORDEIRO, M. G., NETO E. A. L. **Modelos Paramétricos**, 16ª SINAPE 26 a 30/07/04 ABE (Associação Brasileira de Estatística).

CORDEIRO, M. G., **Modelos Lineares Generalizados**, VII SINAPE, Campinas, São Paulo, 1986; **Empréstimos**. 5ª ed. São Paulo: Atlas, 2000.

DIAS FILHO, José Maria. **Gestão tributária na era da responsabilidade fiscal: proposta para otimizar a curva da receita utilizando conceito de semiótica e regressão logística**. 2003, 251 f. Tese (Doutorado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo.

FIGUEIRA, Cleonis Viater **Modelos de Regressão Logística**. Porto Alegre, 2006. Dissertação (mestrado em matemática) Programa de Pós-graduação em matemática Universidade Federal do Rio Grande do Sul.

Filho, F. P. S. **Garantias de operações de crédito**. São Paulo, IBCB, 1990.

GRAYBILL, Franklin A.; IYER Hariharan K. **Regression analysis: Concepts and Applications**. California, Duxbury Press 2006.

GRAYBILL, Franklin A.; IYER, Hariharan K. **Regression analysis: Concepts and Application**. California, Duxbury Press 2006.

GUIMARÃES, L. R.. **Previsão de inadimplência e redes neurais artificiais**. Dissertação de Mestrado. Universidade do estado de Rio de Janeiro.2006 p.1

HAUCK, W. W., DONNER, A. **Wald's test as applied to hypothesis in logit analysis**. Journal of the American Statistical Association. 1977.

HOSMER D; LEMESHOW S. **Applied logistic regression**, Second Edition. New York: John Wiley & Sons, 2000.

HOSMER, David W.; LEMESHOW, Stanley. **Applied logistic Regression**. 2nd. Ed. New York: John Wiley & Sons, 2000.

JANÉ, D. de . A. **Uma introdução ao estudo da lógica fuzzy**.2004 Hórus – Revista de Humanidades e Ciências Sociais Aplicadas, Ourinhos/SP, N° 02, 2004. Disponível:

JOHNSON, Richard A.; WICHERN, Dean W. **Applied multivariate statistical analysis**. 3ª Ed. Nova Jersey, 1992 Prentice Hall.

JOHNSON, Richard A.; WICHERN, Dean W. **Applied multivariate statistical analysis** 3 ed. Nova Jersey, 1992 Prentice Hall, 3 edição 600 p.

JORION, Phillippe. **Value-at-risk: The New Benchmark for Controlling Market Risk**, Chicago: Irwin Professional Publishing, 1997.

KARTALOPOULOS, S. V. **Understanding neural networks and fuzzy logic: basic concepts and applications**. Piscataway: IEEE Press, 1996.

LEWIS, E.M. (1992) **An Introtruction to Credit Scoring**. San Rafael: Fair Isac and Co., Inc.

LIMA, J. (2002) **A Análise Econômico-Financeira de Empresas sob a Ótica da Estatística Multivariada** Dissertação de Mestrado, Curitiba: Universidade Federal do Paraná.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). **Generalized Linear Models**. J. Roy. Statist. Soc. Ser.

NETER, JOHN; KUTNER, M.H; NACHTSHEIM, CRISTOPHER J; WASSERMAN, WILLIAM. **Applied Linear Statistical Models**. 4th Edition. Illinois: Burr Ridger,1996.

PAULA, G. A. (2002) **Modelos de Regressão com Apoio Computacional**, material disponível em <http://www.ime.usp.br/~giapaula/livro.pdf> acesso em 05/12/2004.

PEREIRA, GUSTAVO H. A. **Modelo de Risco de Crédito de Clientes: Uma aplicação a dados reais**. São Paulo, 2004.

PERERA, Luiz Carlos Jacob. **Decisão de Crédito para Grandes Corporações**, Tese (Doutorado em Administração) – FEA/USP, São Paulo: Universidade São Paulo, 1998.

PREARO, Leandro Campi **Uso de Técnicas estatísticas multivariadas em dissertações e teses sobre o comportamento do consumidor: um estudo exploratório**. São Paulo, 2008 100 p. Dissertação (Mestrado) Universidade de São Paulo.

REZENDE. F. C., **Construção de modelos de classificação de Risco de Crédito para empresas brasileiras com base em indicadores contábeis**. Dissertação de Mestrado. IBMEC, São Paulo, 2007.

RYAN, Thomas P. **Modern Regression Methods**. 2^a Ed. John Wiley & Sons, 2009.

RYOSTON, Patrick; SAUERBREI, Willi **Multivariable model-building: A pragmatic approach to regression analysis based on fraction polynomials for continuous variable**. Ed. John Wiley and Sons, 2008.

SANTOS, José Odálio dos, **Análise de crédito: empresas e pessoas físicas**, São Paulo, Atlas, 2000.

SCHRICKEL, Wolfgang Kurt. **Análise de crédito: concessão e gerência de empréstimos**. 3 ed. São Paulo: Atlas, 1997. 331p.

SEMOLINI, R. (2002) **Support Vector Machines, Inferência Transdutiva e o Problema de Classificação**. Dissertação de Mestrado. Departamento de Engenharia Elétrica. Universidade Estadual de Campinas FEEC/UNICAMP.

Shrickel, W. K. **Análise de crédito: Concessão e gerência de empréstimos**, São Paulo, Atlas,1994.

SILVA, JOSÉ P., **Gestão e Análise de Risco de Crédito**, Segunda Edição, São Paulo: Atlas, 1998.

SILVA, JOSÉ P., **Gestão e Análise de Risco de Crédito**, Terceira Edição, São Paulo: Atlas, 2003.

TREVISANI, A. T., GONÇALVES, E. B., D'EMÍDIO, M, HUMES L. L. (2004) **Qualidade de Dados – Desafio crítico para o Sucesso do Intelligence**, Itajaí: XVIII Congresso Latino Americano de Estratégia.

VASCONCELLOS, M. S. **Proposta de Método para Análise de Concessões de Crédito a Pessoas Físicas**. São Paulo. 2002.

WEISBERG, Sanford. **Applied Linear Regression**. 2005 3ª Ed. John Wiley and Sons.