

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
MESTRADO EM ADMINISTRAÇÃO

SAMUEL ONGARATTO

VALOR DO CLIENTE, INADIMPLÊNCIA E ASSIMETRIA DE FLUXO DE CAIXA

Orientador: Prof. Dr. Guilherme Liberali Neto
Co-Orientador: Prof. Dr. Edar da Silva Añaña

São Leopoldo, 2010

SAMUEL ONGARATTO

VALOR DO CLIENTE, INADIMPLÊNCIA E ASSIMETRIA DE FLUXO DE CAIXA

Dissertação apresentada ao Programa de Pós-Graduação em Administração da Universidade do Vale do Rio dos Sinos – UNISINOS como requisito parcial para a obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Guilherme Liberali Neto
Co-Orientador: Prof. Dr. Edar da Silva Añaña

São Leopoldo, 2010

FOLHA DE APROVAÇÃO

Samuel Ongaratto

Título: Valor do Cliente, Inadimplência e Assimetria de Fluxo de Caixa

Dissertação apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre em Administração.

Aprovado em 20 de agosto de 2010.

BANCA EXAMINADORA

Prof. Dr. Lélis Balestrin Espartel - PUCRS

Prof. Dr. Ivan Lapuente Garrido - UNISINOS

Prof. Dr. Celso Augusto de Matos - UNISINOS

Prof. Dr. Guilherme Liberali Neto (Orientador)

Prof. Dr. Edar da Silva Añaña (Co-Orientador)

Visto e permitida a impressão

São Leopoldo,

Prof^ª. Dr^ª. Yeda Swirski de Souza
Coordenador Executivo PPG em
Administração

*“A mente que se abre a uma nova ideia
jamais voltará a seu tamanho original”.*
(Albert Einstein)

AGRADECIMENTOS

Gostaria de agradecer a todos aqueles que contribuíram para que mais esta etapa pudesse ser concluída. Obrigado, pai e mãe, amo vocês. Obrigado a Rose e Isadora, pelo amor e alegria e a Sabrina, pelo companheirismo e amizade.

Obrigado especial ao Liberali, pelas discussões, provocações e desafios que me guiaram durante esta caminhada; ao Edar, pelo incentivo e encorajamento para continuar em frente. Gostaria de agradecer ao Luis Felipe Riehs Camargo, pelo apoio e amizade incondicional. Obrigado ao Funck, pela inspiração no mundo acadêmico.

Quero agradecer a toda minha família e pelo Grupo Ongaratto, que me apoiaram. E aos amigos que fiz durante o mestrado. Gostaria de agradecer ao MOSES grupo de pesquisa, pelas discussões e ideias que ajudaram na realização deste trabalho.

RESUMO

O objetivo deste estudo é ajudar as empresas na tomada de decisão com relação à base de clientes e ajuste de suas promoções de mercado num contexto de risco de crédito comercial. O resultado da pesquisa é a proposição de um novo modelo. Como contribuições teóricas, pode ser citado o desenvolvimento de um modelo de risco de crédito capaz de estimar um risco de inadimplência para cada pagamento efetuado por um cliente. Isso difere dos modelos encontrados na literatura, que estimam apenas um risco para cada cliente. Outra contribuição é o desenvolvimento de um modelo baseado na métrica de *Customer Lifetime Value* com componentes inéditos (assimetria entre prazos de pagamento e recebimento e risco de crédito). Esta pesquisa é dividida em três fases distintas: uma fase exploratória, resultado de uma pesquisa realizada na literatura em busca de conceitos e elementos alinhados ao objetivo; a segunda é a proposição do método e do modelo propriamente ditos, e a terceira e última fase foi a aplicação do modelo a um estudo de campo, o qual utilizou dados de 14.259 faturas e 229 clientes. Os dados são de dezembro de 2007 a agosto de 2009. O modelo de risco de crédito integrado ao modelo proposto classifica as faturas pagas com 75,52% de assertividade média. Os resultados do estudo de campo ajudaram a empresa estudada a realizar uma série de mudanças na sua base de clientes, resultando com essas medidas num ganho estimado de mais de R\$ 2,5 milhões para 2010.

Palavras-chave: *Customer Lifetime Value*, Risco de Crédito, Inadimplência.

ABSTRACT

The objective of this research is to help companies in decision making regarding the customer base and adjust their marketing promotions in the context of commercial credit risk. The result of this research is to propose a new model. As theoretical contributions can be mentioned the development of a model of credit risk can estimate a default risk for each payment made by a client. This differs from the other models in the literature that estimate only one risk for each client. Another contribution is the development of a model based on metrics of Customer Lifetime Value with components unpublished (asymmetry between receiving and payment terms and credit risk). This research is divided into three distinct phases: an exploratory phase, where it performed a literature review in search of concepts and elements aligned to the goal. The second phase is the proposition of the method and the model itself. The third and final phase was the implementation of the model to a field study. The field study used data from 14 259 bills and 229 customers. Data are from December 2007 to August 2009. The model of credit risk built into the proposed model classifies invoices paid on average 75.52% of assertiveness. The results of the field helped the company studied conducting a series of changes in its customer base. Changes made resulting in an estimated gain of more than R\$ 2.5 million in 2010.

Keywords: Customer Lifetime Value, Credit risk, Default.

LISTA DE FIGURAS

Figura 1. Classificação dos modelos de CLV	31
Figura 2. Fluxo de caixa sem assimetria entre prazos de pagamento e recebimento	38
Figura 3. Fluxo de caixa com assimetria entre prazos de pagamento e recebimento.....	38
Figura 4. Curva logística	45
Figura 5. Modelo conceitual da pesquisa	47
Figura 6. Método da pesquisa.....	48
Figura 7. Venda com ordem de frete	51
Figura 8. Localização dos postos de combustíveis.....	52
Figura 9. Curva ROC.....	74
Figura 10. Comparação de modelos com a curva ROC	75
Figura 11. Curva ROC para o modelo de risco de inadimplência.....	76
Figura 11.1. Comparação de modelos utilizando a curva ROC	80
Figura 12. Evolução do risco de inadimplência estimado para o cliente "605e844"	81
Figura 13. Curvas dos valores dos termos do CLV	83
Figura 14. Relação do valor do cliente com o risco de inadimplência.....	86
Figura 15. Estatística descritiva.....	101
Figura 16. Estatística descritiva.....	102
Figura 17. Estatística descritiva.....	103
Figura 18. Estatística descritiva.....	104
Figura 19. Estatística descritiva.....	106
Figura 20. Evolução do risco de inadimplência estimado para o cliente "49e1568"	107
Figura 21. Evolução do risco de inadimplência estimado para o cliente "605e844"	108
Figura 22. Evolução do risco de inadimplência estimado para o cliente "589"	108
Figura 23. Evolução do risco de inadimplência estimado para o cliente "245"	108
Figura 24. Evolução do risco de inadimplência estimado para o cliente "145"	109
Figura 25. Evolução do risco de inadimplência estimado para o cliente "144"	109
Figura 26. Evolução do risco de inadimplência estimado para o cliente "594"	109

LISTA DE QUADROS

Quadro 1. Matriz de correlação.....	64
Quadro 2. Configurações testadas na calibração e projeção	69
Quadro 3. Resultados dos modelos	70
Quadro 4. Resultados dos modelos	72
Quadro 4.1. Conjunto de respostas possíveis.....	73
Quadro 4.2. Resultado Curva ROC.....	76
Quadro 4.3. Comparação de modelos utilizando a curva ROC	79
Quadro 5. Procedimento para integração do modelo de risco de inadimplência ao modelo de CLV	82
Quadro 6. Resultado do modelo proposto.....	85
Quadro 7. Mudanças ocorridas na empresa estudada	90
Quadro 8. Estatística descritiva.....	101
Quadro 9. Estatística descritiva.....	102
Quadro 10. Estatística descritiva.....	103
Quadro 11. Estatística descritiva.....	104
Quadro 12. Estatística descritiva.....	105

LISTA DE TABELAS

Tabela 1. Trabalhos sobre risco de crédito	24
Tabela 2. Resultados dos modelos que utilizaram apenas um tipo de dados	29
Tabela 3. Resultados dos modelos que utilizaram dados combinados	30
Tabela 4. Faturamento por cliente de dezembro de 2007 a agosto de 2009.....	53
Tabela 5. Coleta de dados.....	55
Tabela 6. Alterações realizadas na base de dados	57
Tabela 7. Variáveis preditoras testadas	61
Tabela 8. Resultados da estimação	62
Tabela 9. Resultados da estimação	63
Tabela 10. Resultados da estimação	64
Tabela 11. Resultados da estimação	64
Tabela 12. Classificação do coeficiente de correlação	65
Tabela 13. Resultados do teste FIV	66
Tabela 14. Número de faturas adimplentes e inadimplentes	70
Tabela 15. Divisão da base	71
Tabela 15.1. Pontos de corte na Curva ROC.....	77
Tabela 16. Média do risco de inadimplência para os clientes	80
Tabela 17. Média do risco de inadimplência e número de faturas emitidas por clientes	80
Tabela 18. Preço de compra do diesel	83
Tabela 19. Relação do valor do cliente com o risco de inadimplência	85
Tabela 20. Ranking realizado pela empresa estudada	88
Tabela 21. Ranking sugerido pela pesquisa.....	89
Tabela 22. Série temporal de venda de litros de diesel para os principais clientes da empresa	100
Tabela 23. Estatística descritiva	107
Tabela 24. Pontos de corte para os modelos.....	110

LISTA DE ABREVIATURAS

CLV - *Customer Lifetime Value*

CE – *Customer Equity*

ROC – *Receiver Operating Characteristic*

KM – Quilômetro

VPL – Valor Presente Líquido

CV – *Customer Value*

AC – Ativo Circulante

PC – Passivo Circulante

PL – Patrimônio Líquido

ELP – Exigível a Longo Prazo

LL – Lucro Líquido

RLF – Realizável a Longo Prazo

IF – Fator de Insolvência

CRM – *Customer Relationship Management*

SUMÁRIO

INTRODUÇÃO	14
1. OBJETIVO DO TRABALHO	17
1.1. JUSTIFICATIVA DA PESQUISA	17
1.2. ESTRUTURA DA DISSERTAÇÃO	18
2. REFERENCIAL TEÓRICO	19
2.1. <i>CUSTOMER LIFETIME VALUE</i> (CLV)	19
2.2. MODELOS DE PREVISÃO DE INADIMPLÊNCIA.....	23
2.3. <i>CUSTOMER LIFETIME VALUE</i> NO CONTEXTO DE INADIMPLÊNCIA.....	30
2.4. MODELO PROPOSTO.....	36
2.4.1. Regressão Logística	44
3. MÉTODO	48
3.1. DEFINIÇÃO DA AMOSTRA	49
3.1.1. O Setor	50
3.1.2. A Empresa	52
3.1.3. Período de análise e número de observações	54
3.2. COLETA E TRATAMENTO DOS DADOS.....	56
4. ANÁLISE DOS DADOS	60
4.1. MODELO DE INADIMPLÊNCIA	60
4.1.1. Variáveis explicativas ou Regressoras	60
4.1.2. Calibração e projeção	62
4.2. MODELO PROPOSTO.....	81
5. CONSIDERAÇÕES FINAIS	87
5.1. IMPLICAÇÕES GERENCIAIS.....	88
5.2. LIMITAÇÕES E TRABALHOS FUTUROS	92
REFERÊNCIAS BIBLIOGRÁFICAS	96
APÊNDICE A. SÉRIE MENSAL DE COMPRA DE LITROS DE DIESEL POR CLIENTE	100
APÊNDICE B. ESTATÍSTICA DESCRITIVA	101
APÊNDICE C. ESTATÍSTICA DESCRITIVA DO RISCO ESTIMADO DOS CLIENTES ESTUDADOS	107
APÊNDICE D. RESULTADOS CURVA ROC	110

INTRODUÇÃO

Este trabalho une dois importantes campos de conhecimento na literatura administrativa: marketing e finanças. O objetivo dessa união é ajudar as empresas na tomada de decisão com relação à base de clientes e ajuste de suas promoções de mercado num contexto de risco de crédito comercial. O resultado da pesquisa é a proposição de um novo modelo baseado na métrica de *Customer Lifetime Value* (CLV) e sua posterior aplicação a um contexto real de estudo de campo.

Como contribuições teóricas na área de marketing podem ser citados o desenvolvimento de um modelo baseado na métrica de *Customer Lifetime Value* com componentes inéditos (assimetria entre prazos de pagamento e recebimento e risco de crédito) e a aplicação deste ao estudo de caso real. Diversos trabalhos na literatura propõem inúmeros modelos de CLV, porém poucos demonstram sua aplicação a um contexto real (KUMAR, 2008).

Dentre as contribuições para a área de finanças, pode ser destacado o desenvolvimento de um modelo de risco de crédito capaz de detectar problemas no pagamento dos clientes, sendo este também aplicável ao contexto de crédito comercial. A literatura possui considerável número de trabalhos que servem para prever a falência de possíveis aspirantes a crédito. Embora seja importante saber se um cliente irá ou não tornar-se insolvente, é importante detectar problemas nos pagamentos, até porque uma empresa pode não estar numa situação de insolvência, mas, mesmo assim, ser incapaz de honrar com seus compromissos. Além disso, grande parte dos modelos de previsão de inadimplência utiliza dados de análise de balanço. Esses dados são, muitas vezes, inacessíveis e pouco utilizados em um contexto de crédito comercial. Ainda, o modelo de risco de crédito desenvolvido nesta pesquisa é capaz de estimar um risco de inadimplência para cada pagamento efetuado por um cliente. Isso difere dos modelos encontrados na literatura, que estimam um risco para cada cliente. Essa abordagem contempla uma nova perspectiva: a de que o risco de inadimplência de um determinado cliente varia no tempo (curto prazo).

Com relação ao método, esta pesquisa é dividida em três fases distintas: uma fase exploratória, resultado de uma pesquisa realizada na literatura em busca de conceitos e elementos alinhados ao objetivo; a segunda é a proposição do método e do modelo propriamente ditos, e a terceira fase foi a aplicação do modelo a um estudo de campo em um posto de combustível de rodovia, que trabalha com venda de combustíveis a prazo (crédito

comercial). Este estudo utilizou uma extensa base de dados com 14.259 faturas e 229 clientes. O horizonte de dados coletados é de dezembro de 2007 a agosto de 2009. Os resultados do estudo de campo ajudaram a empresa estudada a realizar uma série de mudanças na sua base de clientes, resultando com isso em um ganho estimado de mais de R\$ 2,5 milhões para 2010.

Nos resultados, são evidentes os casos de relações comerciais prejudiciais à empresa estudada, e a existência destas relações é resultado de pressões do mercado e necessitam de uma maior aproximação com seus clientes. Cada vez mais, a escolha adequada dos clientes-alvo e uma construção de relações de longo prazo podem ser fatores determinantes para o sucesso das empresas. Mesmo depois da definição dos clientes-alvo, manter um relacionamento de longo prazo com eles passa por adequar individualmente às políticas de promoção de vendas o preço, os prazos de pagamentos, entre outros (RUST, ZEITHAML e LEMON, 2001). Nesse contexto, o preço é um importante atributo e é constantemente negociado pelos consumidores. As empresas utilizam descontos, prazos de pagamentos e, em alguns casos, os dois ao mesmo tempo no intuito de aumentar seu faturamento. Apesar de a utilização dos artifícios relacionados ao preço funcionar como fator de atração de clientes, a competição exclusiva por preço não garante futuras contratações ou a fidelidade deles, pois essa é uma prática destrutiva em alguns segmentos (HAYES *et al.*, 2008).

Outra questão associada ao preço são os descontos concedidos nas vendas, cujos percentuais relativamente pequenos de descontos impactam significativamente na lucratividade de uma empresa (ATKIN, POCOOCK e TAYLOR, 1988). Uma pesquisa realizada pela McKinsey em 2.400 empresas demonstrou que a redução de 1% nos custos fixos causa um impacto positivo na lucratividade em 2,3%; um aumento de 1% no volume de vendas eleva a lucratividade em 3,3%; a redução de 1% nos custos variáveis resulta em um aumento de 7,8% nos lucros e, finalmente, um aumento de apenas 1% nos preços eleva o lucro em 11% (RAJU e ZHANG, 2006). Dentre os dados citados na pesquisa realizada pela McKinsey, observa-se que o fator que causa maior impacto nos lucros é o preço e, conseqüentemente, os descontos concedidos.

Assim como os descontos, a venda a prazo também traz consigo fatores complicadores que causam impactos negativos no lucro. Um dos quais é o risco de crédito ou inadimplência. O risco de crédito é a expectativa do recebimento de uma quantia em capital, num determinado espaço de tempo, então o risco de crédito é, na realidade, a chance de essa expectativa não se realizar (CAOUILLE, ALTMAN e NARAYANAN, 1999). Nas vendas a prazo, quanto maior o prazo cedido, maior o risco de crédito e, conseqüentemente, maiores serão as perdas financeiras, pois mais tempo estaria disponível para o cliente se tornar

insolvente (GITMAN, 2008). Somado a isso, o Brasil possui uma taxa de juros em torno de 10,25% a.a. (BACEN, 2010), o que acaba tornando o preço do capital mais alto, onerando, assim, os custos caso se concedam longos prazos aos clientes (perdas financeiras em função do prazo).

Por um lado, a competitividade força as empresas a concederem descontos e prazos de forma a não comprometerem seu faturamento e, por outro, descontos e prazos mal dimensionados causam perdas financeiras. Esta dissertação busca responder à seguinte questão de pesquisa: **como mensurar as relações comerciais com os clientes, em um ambiente de risco de inadimplência, para ajudar a empresa a adequar suas políticas de negociação com os clientes?**

1. OBJETIVO DO TRABALHO

Este estudo tem como objetivo geral **mensurar as relações comerciais com os clientes através da proposição de um modelo que auxilie a empresa a decidir sobre os descontos e prazos de recebimento para os clientes**. A fim de alcançar o objetivo geral, será necessário o alcance dos seguintes objetivos específicos:

- mensurar o risco de inadimplência dos clientes da empresa estudada com a técnica de regressão logística;
- mensurar o *customer lifetime value* dos clientes da empresa estudada.

1.1. JUSTIFICATIVA DA PESQUISA

Atualmente, o cliente é considerado o principal ativo das empresas, sendo ele que adiciona valor à organização. A tarefa de escolher quais clientes focar é hoje alvo de muitas discussões na literatura (RUST, ZEITHAML e LEMON, 2001).

Este trabalho é justificado porque discute o valor do cliente em um contexto de crédito comercial. O trabalho relaciona dois assuntos importantes: o risco de inadimplência e o valor do cliente (CLV), e mostra, através de um estudo de campo, uma proposta para mensurar o valor do cliente, utilizando uma nova sugestão de modelo. Apesar de inúmeros artigos discutirem sobre como mensurar o valor do cliente, são poucos os trabalhos que demonstram como isso funciona. Os estudos de campo são importantes, uma vez que podem estabelecer a validade externa dos modelos de CLV e suas aplicações (KUMAR *et al.* 2008).

Do ponto de vista gerencial, este trabalho ajudou a empresa estudada a entender melhor suas relações com os clientes e serviu como base para uma série de mudanças realizadas nas políticas promocionais. O modelo ajudou a empresa na tomada de decisão com relação à escolha de clientes-alvo, segmentação de clientes, dimensionamento de descontos concedidos, ajustes de prazos de pagamento e criação de políticas de risco de crédito. Comparada a decisões tomadas intuitivamente ou baseadas em processos subjetivos, a criação de um método e de um modelo matemático deu embasamento e robustez ao processo de decisão da empresa.

A utilização do CLV permitiu elaborar um modelo para mensurar as relações comerciais da empresa com seus clientes. Essas informações se mostraram úteis para quantificar as decisões com relação à política de preços adotada. É importante a utilização de informações sobre os consumidores na elaboração de estratégias e no auxílio à tomada de decisão (HAUSER, 1993), já que o cliente é a principal fonte de geração de valor de uma empresa (GUPTA e ZEITHAML, 2006).

1.2. ESTRUTURA DA DISSERTAÇÃO

Esta pesquisa está dividida em cinco capítulos. O primeiro capítulo é o objetivo do trabalho e a justificativa. O segundo é o referencial teórico utilizado na pesquisa. O terceiro discute sobre os métodos utilizados. O quarto capítulo traz uma análise dos resultados, e o quinto contém as considerações finais da pesquisa.

2. REFERENCIAL TEÓRICO

Neste capítulo serão abordados o CLV, os modelos de previsão de inadimplência e as diferentes alternativas estudadas para solucionar o problema desta pesquisa.

2.1. *CUSTOMER LIFETIME VALUE* (CLV)

O CLV tem como objetivo mensurar o valor de um cliente durante o seu período de relacionamento com a empresa. Isso é feito através do Valor Presente Líquido (VPL) das receitas proporcionadas por um cliente durante seu tempo de relação com a empresa (BERGER e NASR, 1998; GUPTA e ZEITHAML, 2006; VILLANUEVA e HANSSENS, 2007).

Para Rust, Zeithaml e Lemon (2001), o Valor do Cliente é um dos componentes mais importantes que compõem o valor da empresa, ainda que este valor não seja o mesmo que o seu patrimônio (ativos, imobilizados, etc.), pois são os clientes que provêm a fonte de recursos futuros necessários para a sua existência. Outro ponto é que mensurar o valor do cliente permite às empresas alocarem recursos de marketing de forma mais eficiente (AERON, 2008; KUMAR *et al.* 2008). Este valor não deve apenas ser percebido em termos de lucratividade proporcionada pelos clientes, mas também na constância do fluxo líquido descontado destas contribuições, as quais podem ser representadas por um fluxo de caixa líquido descontado (PAIVA, 2004). Em finanças, o valor da empresa é dado também pela sua capacidade de gerar fluxo de caixa (ROSS, WESTERFIELD e JAFFE 2008).

O CLV é um modelo matemático que ajuda a empresa a entender e avaliar de maneira sistemática as suas relações com seus clientes (JAIN e SINGH, 2002). Clientes selecionados através do uso do CLV podem prover mais ganhos futuros que os selecionados através de outros indicadores (VENKATESAN e KUMAR, 2004). O termo CLV é bem aceito tanto na academia quanto nas empresas, por outro lado, destaca-se que empresas e indústrias já praticavam seus conceitos há mais tempo (VILLANUEVA e HANSSENS, 2007).

Modelos de CLV podem se diferenciar pelas quantidades e tipos de pressupostos envolvidos neles, e pelo objetivo para o qual ele foi construído. Por este motivo existem diferentes modelos de CLV na literatura. Abaixo segue a seguinte estrutura de um deles (VENKATESAN e KUMAR, 2004):

$$CLV_j = \sum_{t=1}^n \frac{(r_{j,t} - c_{j,t})}{(1+i)^t} \quad (1)$$

Portanto:

$r_{j,t}$ = receitas com o cliente j no período t ;

$c_{j,t}$ = custos com o cliente j no período t ;

j = índice do cliente;

t = índice do tempo;

n = horizonte de previsão;

i = taxa de desconto.

Geralmente, os modelos de CLV seguem uma estrutura básica, porém são ajustados devido a fatores existentes no contexto da empresa. Em situações contratuais, os gestores estão interessados na predição da retenção do cliente, ou na probabilidade da sua permanência nos períodos futuros. Esse tipo de situação é caracterizado como “*lost-for-good*” (RUST, ZEITHAMAL e LEMAN, 2004; DWYER, 1997). Entretanto, em ambientes não-contratuais, o foco é na predição da demanda futura, já que não se possuem garantias de que o cliente irá retornar para efetuar a recompra (VENKATESAN e KUMAR, 2004). Esse segundo cenário é chamado de “*always-a-share*” (RUST, ZEITHAMAL e LEMAN, 2004; DWYER, 1997).

Venkatesan e Kumar (2004) avaliam o CLV como uma métrica para selecionar clientes mais lucrativos. Eles analisam diferentes métricas (CRM) e sinalizam o fato de as empresas que utilizam o CLV para selecionar seus clientes terem retornos melhores em longo prazo se comparadas às que utilizam outras técnicas. Além disso, existem ganhos potenciais para empresas que utilizam a métrica como balizador para redirecionar os recursos e os esforços da organização. Os autores desenvolvem um modelo para mensurar o valor de CLV para um ambiente “*always-a-share*”, sendo:

$$CLV_j = \sum_{y=1}^{T_j} \frac{CM_{j,y}}{(1+i)^{y/f_j}} - \sum_{l=1}^n \frac{\sum_m c_{j,m,l} x_{j,m,l}}{(1+i)^{l-1}} \quad (2)$$

Assim:

CLV_j = valor do cliente j ;

$CM_{j,y}$ = margem de contribuição prevista para o cliente j (obtida através de um modelo de margem de contribuição) é mensurado em dólar;

i = taxa de desconto;

$c_{j,m,l}$ = custos de marketing unitário para o cliente j , no canal m e no ano l ;

$x_{j,m,l}$ = número de contratos para o cliente j no canal m e no ano l ;

f_j = frequência de vendas prevista para o cliente j ;

n = número de anos previstos;

T_j = número de compras previstas para o cliente j até o fim do período de mensuração.

Além de mensurar o CLV para cada cliente, o objetivo dos autores é alocar recursos de forma que o maximize. O modelo de frequência de compras e de margem de contribuição dos clientes é uma função das variáveis dos recursos de marketing encontradas no canal de contato. O objetivo foi o de identificar quais alocações de recursos são mais adequadas para cada cliente através dos vários canais de comunicação de forma a maximizar o CLV (VENKATESAN e KUMAR, 2004).

Mais tarde, Kumar *et al.* (2008) desenvolvem um modelo para a IBM com o objetivo de (a) escolher clientes-alvo, (b) determinar as quantidades de recursos a serem alocadas para os clientes selecionados e (c) selecionar clientes com potenciais lucros futuros. O trabalho foi conduzido com um estudo de campo na IBM e uma vasta base com 2,5 milhões de observações (72 meses para cada um dos 35.131 clientes).

O modelo de Kumar *et al.* (2008) utilizou uma abordagem *always-a-share*, porque esta é mais adequada a contextos não-contratuais (Reinartz and Kumar 2000, Rust *et al.* 2004, Venkatesan and Kumar 2004). O horizonte de cálculo do CLV foi de 3 anos (36 meses), sendo este um prazo considerado suficiente pela IBM devido ao seu tipo de mercado. O modelo utilizado por Kumar *et al.* (2008) foi:

$$CLV_i = \sum_{j=T+1}^{T+36} \frac{p(Buy_{ij} = 1) \cdot \hat{C}M_{ij}}{(1+r)^{j-T}} - \frac{\hat{M}T_{ij} \cdot \overline{MC}}{(1+r)^{j-T}}, \quad (3)$$

Então:

CLV_i = valor do cliente i ;

$p(Buy_{ij})$ = probabilidade prevista de o cliente i comprar no período j ;

$\hat{C}M_{ij}$ = margem de contribuição prevista pelo cliente i , no período j ;

$\hat{M}T_{ij}$ = número previsto de contatos (ou toques) direcionados ao cliente i , no período j ;

\overline{MC} = custo médio para um contato (ou toque);

j = período, em meses;

T = período final de cálculo;

r = fator de desconto (0,0125, 15% taxa anual).

O modelo de Kumar *et al.* (2008) é utilizado para ranquear os clientes e também serve como critério para efetuar os contatos com eles e assim ajudar a IBM a realocar os recursos de marketing. Os contatos com os clientes são realizados através de e-mail eletrônico, televendas, catálogo e e-mail. Os resultados do estudo de campo são encorajadores: a realocação de recursos baseados no modelo, para aproximadamente 14% dos clientes (*midmarket*), mostra um aumento de aproximadamente \$ 20 milhões. Este aumento é obtido sem nenhum aumento nos níveis de investimento dos recursos de mercado.

Gupta e Zeithamal (2006) seguem a fórmula básica de CLV, adotando uma abordagem do fluxo de caixa descontado. O modelo é aplicado aos novos clientes, uma vez que possam ser identificados os custos de aquisição (AC), sendo:

$$CLV = \sum_{t=1}^T \frac{(r_t - c_t)z_t}{(1+i)^t} - AC \quad (4)$$

Assim:

r_t = receita do cliente no tempo t ;

c_t = custos do cliente no tempo t ;

z_t = probabilidade de o cliente repetir a compra no tempo t ;

i = taxa de desconto;

T = horizonte de tempo do CLV;

AC = custos de aquisição.

Os modelos de CLV representam uma alternativa para os objetivos desta pesquisa. Os modelos com fluxos de caixa descontado (GUPTA e LEHMANN, 2003; GUPTA e ZEITHAML, 2006; VENKATESAN e KUMAR, 2004) mensuram o valor do cliente com o total de contribuições monetárias para a empresa.

Conhecer o CLV ajuda as empresas a ranquear os clientes e melhor definir políticas promocionais. O CLV aparece como alternativa melhor que as outras, porque o valor é mensurado olhando-se para o futuro e não apenas para o passado (GUPTA e LEHMANN, 2003; GUPTA e ZEITHAML, 2006; VENKATESAN e KUMAR, 2004, KUMAR *et al.* 2008). Porém nenhum dos modelos discutidos aborda a questão da inadimplência. A seguir serão vistos modelos que fazem isso.

2.2. MODELOS DE PREVISÃO DE INADIMPLÊNCIA

Muitos estudos têm sido realizados sobre risco de crédito com foco na falência de empresas (ALTMAN, BAIDYIA e DIAS, 1979; BEAVER 1966; KANITZ, 1974; SANVICENT e MINARDI, 1998). Na maioria, o objetivo destes estudos é desenvolver modelos com alto poder discriminante para ajudar instituições financeiras a decidir a conceder ou não crédito para clientes. Para esses estudos, a previsão da insolvência de empresas devedoras tem sido a principal preocupação. No Brasil, a taxa de mortalidade das micro e pequenas empresas chega a 51%, até 2 anos de idade, dados de 2002 (SEBRAE, 2009).

Embora a insolvência, definida como a falência da empresa, ou ainda, uma condição binária (a empresa está ou não falida) seja preocupação para instituições de crédito, pouco se tem feito com relação ao risco de crédito. A insolvência é apenas um resultante da evolução do risco de crédito (CAOUILLE, ALTMAN e NARAYANAN, 1999). E ainda são poucos os estudos que buscam prever a inadimplência (EIFERT, 2003).

A pesquisa relacionada à previsão de insolvência de empresas iniciou na década de 60 com o trabalho de Beaver (1966). Após, inúmeros trabalhos, alguns dos quais estão na Tabela 1, foram realizados utilizando diferentes técnicas.

Autor/Ano	Técnica Utilizada
BEAVER (1966)	Análise Discriminante Univariada
ALTMAN (1968)	Análise Discriminante Múltipla
KANITZ (1974)	Análise Discriminante Múltipla
ALTMAN, BAIDYIA e DIAS (1979)	Análise Discriminante Múltipla
BACK <i>et al.</i> (1996)	Redes Neurais, Regressão Logística e Análise Discriminante Linear Múltipla
SANVICENT e MINARDI (1998)	Análise Discriminante Múltipla
BERTUCCI <i>et al.</i> (2003)	Análise Discriminante Múltipla e Regressão Logística
EIFERT (2003)	Análise Discriminante Múltipla e Regressão Logística
RIBEIRO (2008)	Regressão logística
WILSON, SUMMERS e HOPE (2000)	Regressão logística

Tabela 1. Trabalhos sobre risco de crédito.

Para Beaver (1966), a empresa é descrita como um reservatório de ativos líquidos. O reservatório é mantido por entradas e saídas de fundos, sendo necessário que a empresa mantenha um nível mínimo para proteger-se das variações do mercado. A partir desta visão, o autor considera como insolventes empresas em que o reservatório foi exaurido, não podendo honrar com seus compromissos.

Beaver (1966) comparou 29 índices com uma escala de tempo de 1 até 5 anos antes da insolvência. Foi utilizada uma amostra de 159 empresas (50% solventes e 50% insolventes). O índice de maior poder discriminante foi o fluxo de caixa dividido sobre a dívida total. Este índice obteve um percentual de acerto de 87% para um ano antes da falência e 78% para cinco anos antes.

Embora Beaver tenha alcançado bons resultados preditivos, a análise discriminante linear é uma técnica univariada, a qual analisa o poder preditivo das variáveis separadamente, não considerando a interdependência entre os indicadores. No intuito de sanar esta deficiência, Altman (1968) utiliza a análise discriminante múltipla e desenvolve um modelo conhecido por *Z-score*. Foram utilizados dados de 66 empresas (33 solventes e 33 insolventes) com características semelhantes (indústria, tamanho, entre outros). Altman (1968) utiliza 22 indicadores, sendo o modelo final:

$$Z = 0,012X_1 + 0,014X_2 + 0,033X_3 + 0,006X_4 + 0,999X_5 \quad (5)$$

Dessa forma:

$$X_1 = AC - PC / \text{ativo total};$$

$$X_2 = \text{lucros acumulados} / \text{ativo total};$$

$$X_3 = \text{lucro antes dos juros e impostos} / \text{ativo total};$$

$X_4 = \text{valor de mercado do PL} / (\text{PC} + \text{ELP});$

$X_5 = \text{vendas} / \text{ativo total}.$

O Fator de Insolvência (IF) ou o Termômetro de Kanitz foi desenvolvido por Kanitz (1974) e foi o primeiro trabalho sobre modelos de previsão de insolvência desenvolvido no Brasil. Utilizando análise discriminante linear e dados de contabilidade, Kanitz (1974) desenvolveu um indicador que permite avaliar se a empresa se encontra ou não em uma faixa de perigo de insolvência. O resultado varia de 7 a -7, sendo classificado como: 7 a 0, a empresa está numa situação de solvência; de 0 a -3, está em uma situação de problemas financeiros; e de -3 a -7, ela está numa situação de pré-insolvência. Para esta classificação, foi elaborado o seguinte modelo:

$$IF = 0,05X_1 + 1,65X_2 + 3,55X_3 - 1,06X_4 - 0,33X_5 \quad (6)$$

Assim:

$X_1 = \text{LL} / \text{PL};$

$X_2 = \text{AC} + \text{RLP} / \text{PC} + \text{ELP};$

$X_3 = \text{AC} - \text{estoques} / \text{PC};$

$X_4 = \text{AC} / \text{PC};$

$X_5 = \text{PC} + \text{ELP} / \text{PL}.$

Altman, Baidyia e Dias (1979) utilizaram uma amostra de 58 empresas brasileiras (23 com problemas financeiros). Com a utilização de análise discriminante, obtiveram a seguinte equação:

$$Z_1 = -1,44 + 4,03X_2 + 2,25X_3 + 0,14X_4 + 0,42X_5 \quad (7)$$

Portanto:

$X_2 = (\text{PL} - \text{capital subscrito}) / \text{ativo total};$

$X_3 = \text{lucro antes dos juros e impostos} / \text{ativo total};$

$X_4 = \text{PL} / (\text{PC} + \text{ELP});$

$X_5 = \text{vendas} / \text{ativo total}$.

O modelo de Altman, Baidyia e Dias (1979) obteve um sucesso de 87% na classificação de empresas até um ano antes dos problemas financeiros. Com dois anos de antecedência, obteve 84,2%, e três anos 77,8%.

Back et al. (1996) utilizou três modelos, cada um com uma técnica diferente, sendo: análise discriminante linear, regressão logística e redes neurais. Utilizando dados de três anos antes da insolvência, o modelo que utilizou redes neurais obteve um melhor desempenho de classificação, seguido pelo modelo que utilizou regressão logística. Por outro lado, o modelo que utilizou análise discriminante obteve uma melhor classificação, utilizando dados com dois anos antes da insolvência.

Sanvicent e Minardi (1998) utilizaram dados da contabilidade, análise discriminante e uma amostra de 92 empresas (50% sadias e 50% concordatárias) com ações negociadas na bolsa de valores no período de 1986 a 1998. Eles elaboraram um modelo que classificou com uma taxa de 80,2% de acerto as empresas com até um ano de antecedência à concordata.

Bertucci *et al.* (2003) definiram inadimplência como atrasos maiores que 180 dias no pagamento. Eles utilizaram uma amostra composta por micro e pequenas empresas que receberam crédito entre 1998 e 2001 do Banco de Desenvolvimento de Minas Gerais. Utilizando análise discriminante e regressão logística, os autores concluíram que, na medida em que as empresas estudadas aplicam seus recursos em investimentos fixos, maiores são as chances de inadimplência.

Eifert (2003) utiliza dados dos três últimos balanços patrimoniais, juntamente com análise discriminante linear e regressão logística. O autor analisou 51 empresas (21 inadimplentes e 30 adimplentes) industriais que receberam crédito num período de 1996 a 1997. Foram elaborados três modelos, utilizando regressão logística, sendo que o modelo que utilizou dados dos três últimos balanços patrimoniais alcançou uma taxa de sucesso de classificação de 100%. Já os modelos que utilizaram dados dos dois últimos e do último balanço patrimonial conseguiram uma taxa de classificação de 90,2% e 88,2%, respectivamente.

Ribeiro (2008) desenvolve um modelo de previsão de inadimplência para uma instituição de ensino superior privada. O contexto é o de concessão de crédito comercial para pessoa física. As variáveis independentes foram obtidas através de um questionário socioeconômico, contendo um total de 59 variáveis. O questionário foi aplicado com a

utilização de um site corporativo da empresa estudada. A amostra utilizada foi de 256 alunos, sendo que, deste total, apenas 34 estavam classificados em uma situação de inadimplência. A definição de inadimplência utilizada foi de atrasos superiores a 45 dias. Esta definição foi adotada segundo critérios da instituição estudada.

No método, a autora utilizou o teste do Qui-quadrado para verificar a significância estatística de as variáveis discriminarem adimplentes e inadimplentes. A pesquisa concluiu que apenas 3 do total de variáveis eram suficientes para classificar com 65,5% os clientes inadimplentes dos adimplentes em uma amostra de validação (RIBEIRO, 2008). O modelo foi:

$$\ln\left(\frac{p}{(1-p)}\right) = -2,457 + 0,616X_{23} + 2,196X_{30} - 1,263X_{57} \quad (8)$$

X_{23} = reprovação em alguma disciplina;

X_{30} = já negociou débitos;

X_{57} = possui cartão de crédito.

Wilson, Summers e Hope (2000) desenvolvem uma série de modelos utilizando regressão logística. Os objetivos dos autores são: (a) prever o comportamento futuro com relação ao pagamento de débitos comerciais e (b) prever futuras insolvências. Segundo eles, este foi o primeiro estudo realizado com o intuito de prever o comportamento futuro de pagamentos, sendo isso especialmente importante para gestores no contexto de crédito comercial.

Os autores utilizaram um farto banco de dados com 7034 empresas, sendo 3133 destas insolventes. A amostra é heterogênea, contendo diversas empresas de diferentes segmentos e tamanhos. Os dados da amostra referente ao comportamento de pagamentos possuem um caráter longitudinal de tempo, compreendendo dezembro de 1990 a novembro de 1992. De acordo com os autores, no ano de 1992 houve uma forte recessão na economia do Reino Unido, contribuindo para que houvesse um número expressivo de insolvência de empresas. No estudo foram utilizados 25% da amostra (1774 empresas) para teste, e 75% (5260 empresas) para estimação do modelo.

Em cada categoria de dados foram realizados testes para identificar os fatores discriminantes. Inicialmente, os autores utilizaram uma amostra reduzida (3200 empresas)

para a construção do modelo. Esta amostra foi selecionada de forma que não houvesse a existência de dados faltantes (*missing values*). A amostra reduzida era composta por 50% de empresas solventes e 50% de empresas insolventes. 2400 empresas da amostra reduzida foram utilizadas para construção do modelo e 800 empresas para teste.

Todos os modelos foram construídos utilizando dados até dezembro de 1991. Os modelos de previsão de insolvência tinham um horizonte de predição de 12 meses, já os modelos de comportamento de pagamento tinham um horizonte de predição de 6 meses.

As variáveis de comportamento no pagamento foram modeladas em dois diferentes aspectos: (a) empresas que atrasam o pagamento em média 15 dias para um ou mais meses (PAYLTE70), e (b) empresas que atrasam o pagamento em média 30 dias para um ou mais meses (PAYLTE50). Na amostra utilizada, cerca de 74% das empresas se encaixavam como empresas que atrasavam em média 15 dias, outras 27% nas que atrasavam em média 30 dias.

Nos modelos utilizados para prever o comportamento futuro, foram utilizados dados de comportamento de pagamento passado e dados da contabilidade. Os dados de contabilidade foram utilizados para dois diferentes modelos: (a) empresas que pagarão com 15 dias de atraso e (b) empresas que pagarão com 30 dias de atraso. Estes modelos obtiveram uma média de assertividade na classificação de 58,65% e 60,72%, respectivamente. Os dados de comportamento de pagamento passados foram igualmente usados para dois modelos: (a) empresas que pagarão com 15 dias de atraso, e (b) empresas que pagarão com 30 dias de atraso. Estes modelos obtiveram bons resultados, ficando com uma média de assertividade na classificação de 84,61% e 87,10%, respectivamente. Destaque para os resultados dos últimos dois modelos, dos quais foram utilizados apenas dados do comportamento de pagamento passado. Os autores construíram também modelos mistos, utilizando dados de comportamento de pagamento passado e dados da contabilidade, mas não houve melhoras significativas na classificação dos modelos. Uma exceção foi a adição da variável idade da empresa (AGERISK0), que melhorou ligeiramente o poder de predição do modelo de 30 dias, não surtindo o mesmo efeito no modelo de 15 dias.

Para os modelos de predição de insolvência, os autores utilizaram dados de três diferentes fontes: dados contábeis, dados de comportamento de pagamento passado e dados não-financeiros (idade da empresa e setor). Primeiramente, os autores investigaram o poder preditivo de cada categoria de variável separadamente. Para isso, elaboraram modelos que utilizassem apenas uma das três categorias de dados (ver Tabela 2). Em todos os modelos, os autores utilizaram variáveis dummy para acomodar casos em que faltavam os dados.

Modelo	Tipo de dados	Assertividade do modelo		
		% Solvente	% Insolvente	% Geral
1	Modelo Base (1991 clientes)	57,7	73,32	65,51
2	Não-financeiros	36,49	86,76	61,63
3	Comp. de pagamento	49,24	80,04	64,64

Tabela 2. Resultados dos modelos que utilizam apenas um tipo de dado.

Fonte: Adaptado de Winson, Summers e Hope (2000).

Nos três casos, os modelos obtiveram um melhor percentual na classificação de empresas insolventes. A classificação média entre empresas sadias e empresas insolventes variou entre 61,63% e 65,51%.

O modelo com dados contábeis (1) utilizou uma especificação linear do log das variáveis. No modelo que utilizou dados não-financeiros (idade da empresa, classificação da indústria), 372 casos não estavam disponíveis à classificação da indústria, para a qual (,) foram utilizadas variáveis dummy, onde se atribuía 1 para empresas do grupo industrial da pergunta e 0 para empresas de outros grupos. Aos 372 casos que não possuíam classificação foi atribuído o valor 0. Com relação à idade, havia a hipótese da existência de uma correlação não-linear da idade da empresa com a insolvência. Utilizou-se a idade e a idade ao quadrado como variáveis. Aqui também foram utilizadas variáveis *dummy*. Foi atribuído 1 para as empresas que possuíam uma idade entre 2 e 9 anos (idade julgada como de maior risco de inadimplência) e 0 para a restante. Para os autores, as empresas mais novas possuem maior propensão ao inadimplemento, porém é necessário algum tempo para a empresa cometer falhas que gerem débitos, os quais conduzam à falência (o que se revelou sendo verdade, pois esta idade foi a variável com maior poder discriminatório do modelo de dados não-financeiro).

O modelo que utilizou dados de comportamento de pagamentos passados esperava que fosse útil para prever a insolvência, pois os autores acreditavam que empresas com problemas de insolvência também tinham dificuldades em pagar suas dívidas. As variáveis utilizadas para este modelo incluíram os 12 pagamentos mensais de 1991 e um número de variáveis sumárias produzidas a partir delas.

Após estimar os modelos utilizando os três tipos de dados separadamente (dados de comportamento de pagamento, não-financeiros e contábeis), os autores estimaram uma série de modelos, aplicando uma combinação destes tipos de dados. Os melhores resultados dos modelos estimados em cada combinação estão descritos na tabela 3.

Modelo	Tipo de dados	Assertividade do modelo		
		% Solvente	% Insolvente	% Geral
1	Base+1990 clientes	60,73	76,27	68,50
2	Base+não-financeiros	57,45	78,82	68,13
3	Base+comportamento	58,59	77,09	67,84
4	Base+1990 clientes+não-financeiros	60,35	78,00	69,18
5	Base+1990 clientes+comportamento	63,01	77,70	70,36
6	Base+não-financeiros+comportamento	61,11	80,55	70,83
7	Base+1990 clientes+não-financeiros+comportamento	63,51	80,24	71,88

Tabela 3. Resultados dos modelos que utilizam dados combinados.

Fonte: Adaptado de Winson, Summers e Hope (2000).

A partir da Tabela 3, são observadas as diferentes sinergias, utilizando diferentes combinações de dados. Outra constatação é que a utilização de dados de comportamento de pagamento passados contribui para melhorar a predição de insolvência de empresas. Isso é observado ao se comparar o poder preditivo do modelo 2 (Base+não-financeiros) com o modelo 6 (Base+não-financeiros+comportamento). A adição de dados de comportamento de pagamento melhorou o poder preditivo médio do modelo de 68,13% para 70,83%, respectivamente.

A inadimplência é definida de diversas maneiras. Eifert (2003) a define como toda a cobrança que é realizada por meio judicial. Wilson, Summers e Hope (2000) a definem em clientes que pagam em média 15 dias atrasados e clientes que pagam em média 30 dias atrasados. Bertucci *et al.* (2003) definem inadimplência como atrasos iguais ou superiores a 180 dias. Embora seja necessário definir inadimplência, outro cuidado a ser tomado é a escolha adequada da técnica que será utilizada para estimá-la.

2.3. CUSTOMER LIFETIME VALUE NO CONTEXTO DE INADIMPLÊNCIA

O contexto em que se pretende estudar é o de risco de crédito comercial. Aqui o recebimento da receita (r) se dá numa data posterior ao ato da venda, existindo assim o risco do não-recebimento. Aeron *et al.* (2008) elaboram um modelo de CLV que incorpora a probabilidade do não-recebimento da receita. O objetivo do modelo é ajudar administradores a tomar decisões com relação à base de clientes para empresas de cartão de crédito. A

crecente competição no mercado de cartão de crédito proveniente de instituições bancárias e não-bancárias eleva os custos de aquisição de novos clientes. Nesse contexto, é importante que a empresa saiba identificar corretamente os clientes-alvo e os mantenha para aumentar seu tempo de vida comercial. Para que ela obtenha sucesso, o conjunto de decisões inicia por selecionar clientes com alto potencial de ganhos futuros e baixo risco de inadimplência.

Depois de selecionar quais são os clientes mais rentáveis, a empresa de cartão de crédito necessita decidir uma série de questões, tais como: o limite do crédito concedido e o preço (taxa) para cada cliente. Além disso, é necessário decidir quem reter e o quanto será gasto na retenção. Por esses motivos, os autores utilizaram o CLV como base para a construção do modelo. Para Aeron *et al.* (2008), os modelos de CLV existentes na literatura podem ser classificados de acordo com a Figura 1.

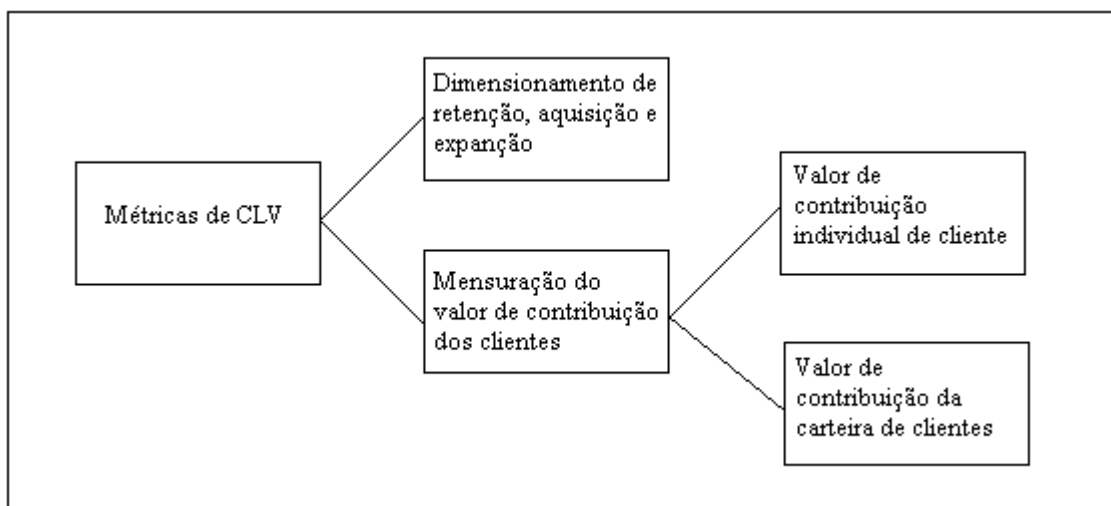


Figura 1. Classificação dos modelos de CLV.
Fonte: Aeron *et al.* (2008).

Para Aeron *et al.* (2008), as métricas existentes na literatura podem ser classificadas em dois grupos: (a) métricas para mensurar valores ótimos de alocação de recursos entre retenção, aquisição e expansão, e (b) métricas que mensuram o valor (entendido aqui como ganho monetário) contribuído por cada cliente para a empresa. A segunda classe de métricas é ainda subdividida em métricas para mensuração da contribuição individual de cada cliente e métricas para mensuração de um conjunto de clientes (carteira, portfólio, entre outros).

Jain e Singh (2002) fazem uma distinção entre os modelos, classificando-os em três grupos: (a) modelos para calcular o CLV; (b) modelos de análise da base de clientes e (c) modelos normativos de CLV. À categoria de modelos para cálculo do CLV são incluídas

especificamente as fórmulas que servem para calcular o CLV e/ou estender este cálculo para obter métodos ideais para alocação de recursos, otimizando o seu resultado.

Dentre o grupo de métricas utilizadas para mensurar o valor (ganho monetário) de cada cliente, os autores levantaram uma série de restrições que impossibilitaram a replicação da métrica para empresas de cartão de crédito, sendo: (a) as métricas não consideram o risco de inadimplência de cada cliente; (b) a variação no rendimento, conforme seu comportamento (compras, retiradas de dinheiro, ficar em débito, entre outros) não é modelado pelas métricas existentes, e (c) o rendimento de um cartão de crédito depende da quantidade de dinheiro pedida e da quantidade devolvida pelo cliente. A quantidade monetária utilizada, a qual pode ser pedida no período (t), é afetada pela quantidade pedida e devolvida no período ($t-1$). Isso ocorre devido à presença de limite de crédito.

Segundo as necessidades específicas discutidas acima, os autores constroem um modelo conceitual para classificar os clientes de empresas de cartão de crédito. Ele simula os diferentes estados que um cliente pode assumir durante seu tempo de vida para a empresa. Os estágios assumidos pelos clientes dependem dos empréstimos realizados e do seu comportamento de pagamento. Os estágios são:

- 1) *Inactive*: significa que o cliente está inativo. Aqui não existem pagamentos oriundos do ciclo passado, nem realização de empréstimos no ciclo atual.
- 2) *Transact*: este estado é definido por um de dois motivos: (a) o cliente está pagando um montante devido do ciclo anterior; ou o cliente está fazendo um empréstimo no ciclo atual.
- 3) *Revolve*: o cliente efetuou um pagamento insuficiente de um empréstimo de um ciclo passado, porém maior que o pagamento mínimo especificado.
- 4) *Delinquent*: o cliente não efetuou o pagamento de um empréstimo ou efetuou um pagamento menor que o mínimo especificado. Aqui há a incidência de taxas adicionais causadas pelo atraso no pagamento.
- 5) *Default*: neste estágio o cliente não efetuou os pagamentos requeridos.

Para determinar estes diferentes estágios, são obedecidas duas regras: (a) cada estágio é definido no final do ciclo de faturamento, dependendo do comportamento de pagamento do cliente, e (b) um cliente pode permanecer no estágio *delinquent* apenas um período, após o qual, e caso ele não efetue o pagamento, será classificado como *default* (no modelo, um período foi compreendido como 0-30 dias, podendo este ser estendido).

A empresa do cartão de crédito pode afetar o comportamento de pagamento e de empréstimo do cliente, alterando as características do cartão, tais como: taxas de juros e limite de crédito, que afetam, por sua vez, o valor de contribuição (m) (ganhos) dos clientes. Quando um cliente é adquirido pela empresa de cartão de crédito, este inicia no estado *inactive*, permanecendo assim até que inicie a utilização do seu cartão.

O valor contribuído pelo cliente é afetado pelo estágio em que este se encontra no período (t). O valor de contribuição é também afetado pelos: (a) custos de manutenção e expedição do banco; (b) taxa cobrada no estágio no qual o cliente é classificado, sendo esta taxa recebida quando ele efetuar o pagamento; (c) taxas adicionais sobre o atraso nos pagamentos; (d) a não-acumulação das taxas, ou seja, elas não são carregadas nos próximos ciclos de empréstimos, e (e) taxa de transação fixa.

Para estimar os rendimentos para cada ciclo, foi necessário estimar dois conjuntos de parâmetros para cada estágio: as probabilidades de um cliente ser classificado em cada um dos estágios, e os montantes associados a cada estágio. Dessa forma, os rendimentos são estimados pela probabilidade de ocorrência do estágio multiplicado ao ganho do empréstimo.

O CV da parcela de empréstimo do estágio *transact* pode ser representado como $ap_t^{Bor} A_t^{bor}$ em que a é a taxa percentual e p_t^{Bor} e A_t^{bor} são a probabilidade e a quantidade emprestada, respectivamente. O rendimento do estágio *revolve* pode ser expressa como $lp_t^{Del} A_t^{Del}$, na qual l são as taxas para pagamento atrasado, p_t^{Del} é a probabilidade de tornar-se *delinquent*, e A_t^{Del} é a quantidade associada ao estágio *delinquent*.

Há um custo de manutenção constante (W) em cada estágio. A instituição bancária incorre custos dos fundos para cobertura das e não pagas. Este custo é descrito como $b(p_t^{Bor} A_t^{Bor} + p_t^{Rev} A_t^{Rev} + p_t^{Del} A_t^{Del})$, em que b é a taxa básica de juros. Um cliente pode fazer empréstimo mesmo estando em um estágio *revolve* ou *delinquent*. A perda de um cliente devido ao *default* é descrita como $p_t^{Def} A_{t-1}^{Def}$, na qual p_t^{Def} é a probabilidade para o *default*, e A_{t-1}^{Del} é o valor do último período (*delinquent*) associado com o estágio *default*.

O CV do cliente da empresa de cartão de crédito no período (t) pode ser expresso como o ganho proveniente ao estágio deste período menos os custos incorridos nesse período. Os rendimentos podem vir de diferentes estágios (*transact*, *revolve*, *delinquent*), e perdas serão provenientes do estágio *default*. Um CV de um determinado estágio num período (t) pode ser colocado como rendimentos para três estágios, sendo: *transact*, *revolve* e *delinquent*,

menos os custos e perdas com manutenção e perdas com o *default*. O CV no final do ciclo (t) pode ser dado por:

$$CV_t = (ap_t^{Bor} A_t^{Bor} + ip_t^{Rev} A_t^{Rev} + lp_t^{Del} A_t^{Del}) - (W + b(p_t^{Bor} A_t^{Bor} + p_t^{Rev} A_t^{Rev} + p_t^{Del} A_t^{Del}) + p_t^{Del} - A_{t-1}^{Del}) \quad (9)$$

ou

$$CV_t = ((a-b)p_t^{Bor} A_t^{Bor} + (i-b)p_t^{Rev} A_t^{Rev} + (l-b)p_t^{Del} A_t^{Del} - p_t^{Def} A_{t-1}^{Del} - W) \quad (10)$$

Existe uma relação entre o montante do estágio *transact*, o do estágio *revolve* e o do estágio *delinquent*. O montante do estágio *revolve* no período (t) será um percentual da quantia emprestada, mais os montantes dos estágios *revolve* e *delinquent* no período ($t-1$), os quais serão:

$$A_t^{Rev} = m_t (A_{t-1}^{Rev} + A_{t-1}^{Bor} + A_{t-1}^{Del}) \quad (11)$$

Sendo que m_t é o valor percentual, devem existir valores de limite superiores e inferiores para m_t . O limite superior é o conjunto de pagamentos mínimos estabelecido pelo banco. Um pagamento que não atinge esta quantidade mínima irá classificar o cliente como *delinquent*. O limite inferior é zero. Se o valor do pagamento for zero no estágio *revolve*, isso significa que não houve giro. O valor de m_t pode variar em cada período. Igualmente para a quantidade do estágio *delinquent* no período (t), será um percentual da quantidade emprestada no período anterior ($t-1$) no estágio *revolve*, assim:

$$A_t^{Del} = n_t (A_{t-1}^{Rev} + A_{t-1}^{Bor}) \quad (12)$$

O n_t é o valor percentual e o limite inferior de n_t é o mesmo limite superior de m_t , pois o pagamento deve ser menor que o mínimo para que o cliente seja classificado como *delinquent*. O limite superior de n_t é 100%. Assim a equação de CV pode ser escrita como:

$$\begin{aligned}
CV_t = & (a-b)p_t^{Bor} A_t^{Bor} + (i-b)m_t p_t^{Rev} (A_{t-1}^{Rev} + A_{t-1}^{Bor} + A_{t-1}^{Del}) \\
& + (l-b)n_t p_t^{Del} (A_{t-1}^{Rev} + A_{t-1}^{Bor}) - p_t^{Def} A_{t-1}^{Del} - W
\end{aligned} \tag{13}$$

Esta equação pode ser reescrita como:

$$\begin{aligned}
CV_t = & (a-b)p_t^{Bor} A_t^{Bor} + (i-b)m_t p_t^{Rev} + ((l-b)n_t p_t^{Del}) A_{t-1}^{Rev} + ((i-b)m_t p_t^{Rev} \\
& + ((l-b)n_t p_t^{Del}) A_{t-1}^{Bor} + ((i-b)m_t p_t^{Del} - p_t^{Def}) A_{t-1}^{Del} - W
\end{aligned} \tag{14}$$

Nas equações 6 e 7, os termos A_{t-1}^{Rev} e A_{t-1}^{Del} podem ser expressos como a quantidade emprestada no processo, que pode terminar com a devolução da quantidade inicialmente pedida.

O propósito desta métrica é o mesmo que a margem $(R_j - C_j)$ proposta na fórmula de CLV de Jain e Singh (2002), em que R_j são as receitas provenientes de um cliente j e C_j são os custos de um cliente j . Para Aeron *et al.* (2008), o CLV pode ser formulado como:

$$CLV = \sum_{t=1}^T CV_t / (1+i)^t \tag{15}$$

O CLV acima não considera as potenciais perdas de oportunidade resultantes do desgaste nas relações futuras (L), uma vez que considera o CV cheio ao longo de todo o horizonte de estimação. Como a perda de receita é inerente ao desgaste, a métrica do CLV deve ser corrigida para:

$$CLV = \sum_{t=1}^T CV_t / (1+i)^t - L \tag{16}$$

Nela o valor da perda por desgaste (L) é computado ao CV a partir do ponto no qual a perda se iniciou. Se o cliente fizer negócios com a empresa durante todo o período, então não há perdas com L . Mas se ele abandonar a relação antes de $(T-t)$ períodos) então haverá um custo de oportunidade devido à perda de receitas esperadas.

A vantagem desta métrica é que os parâmetros de probabilidade de um cliente tornar-se *delinquent* ou *default* podem ser associados às suas características com a utilização de técnicas estatísticas multivariadas. Isso ajuda a empresa a decidir que clientes adquirir e quais

deve reter. O modelo pode ajudar gerentes a decidir os parâmetros do crédito para os clientes (limites, taxas, entre outros), e pode mostrar o quanto a empresa irá ganhar nos períodos futuros, comparando diferentes combinações de parâmetros.

Com relação aos objetivos desta pesquisa, o modelo descrito acima contempla a questão de inadimplência, porém num contexto diferente do estudo de caso desta pesquisa. Além disso, não é necessário que o modelo contemple uma série de outras variáveis de forma a atender os objetivos deste trabalho. Assim foi proposto um novo modelo que será descrito a seguir.

2.4. MODELO PROPOSTO

Este capítulo abordará o modelo matemático proposto e as técnicas que serão utilizadas no estudo de caso. Os conceitos abordados no capítulo anterior dão base para a construção do modelo e das escolhas aqui definidas.

Dentre as métricas estudadas para mensurar o valor (ganho monetário) de cada cliente, as maiores restrições para replicação dos modelos de CLV estudados são: (a) nas vendas a prazo, há a existência de assimetrias entre os prazos de pagamento dos custos e o recebimento das receitas dos clientes, e (b) a maioria das métricas não consideram o risco de inadimplência de cada cliente. Dentre os modelos de CLV estudados, o modelo descrito por Venkatesan e Kumar (2004), com algumas adaptações, aparece como uma possibilidade para mensurar o CLV dos clientes no caso estudado, sendo:

$$CLV_j = \sum_{t=1}^n \frac{(r_{jt} - c_{jt})}{(1+i)^t} \quad (17)$$

CLV_j = o CLV do cliente j ;

j = índice do cliente;

t = índice do período;

r_{jt} = a receita proporcionada no período t , pelo cliente j ;

c_{jt} = o custo do período t , do cliente j ;

i = custo do capital para a empresa.

No caso, trata-se de clientes de uma rede de postos de combustível de rodovia. Neste seguimento, aproximadamente 96% do faturamento da empresa provém da venda de óleo diesel, por este motivo, o CLV dos clientes será mensurado utilizando apenas compras com diesel. Assim, no lugar de utilizar as variáveis “ r ” (receitas geradas pelo cliente) e “ c ” (custos com o cliente), serão utilizadas variáveis relacionadas à quantidade comercializada (q_j) que multiplica o preço de venda (P) e o preço de custo (C), assim o modelo pode ser reescrito da seguinte forma:

$$CLV_j = \sum_{t=1}^T \left(\frac{q_j P - q_j C}{(1+i)^t} \right) \quad (18)$$

Em que:

CLV_j = o CLV do cliente j ;

j = índice do cliente;

t = índice do período;

q_j = a quantidade comercializada pelo cliente j ;

P = o preço unitário de venda do diesel, em reais;

C = o preço unitário de compra do diesel, em reais;

T = horizonte de previsão do CLV;

i = custo do capital para a empresa.

A empresa estudada compra o diesel do fornecedor com um prazo médio de pagamento de quinze dias. Os clientes que trabalham com ordem de frete possuem um prazo de faturamento, que é definido individualmente. O prazo de recebimento pode variar de 3 a 60 dias. Esta situação é definida aqui como assimetria entre os prazos de recebimento e os de pagamento.

Nos modelos de CLV estudados, os prazos de pagamento e recebimento de uma transação com um cliente são simétricos. Da mesma forma, estes modelos não abordam questões referentes à venda ou compra a prazo. Para estes modelos, um fluxo financeiro de uma relação da empresa com um cliente é representado da seguinte maneira:

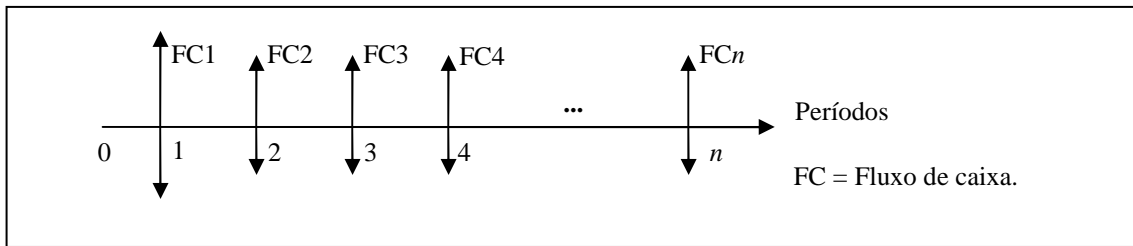


Figura 2. Fluxo de caixa sem assimetria entre prazos de pagamento e recebimento.

A Figura 2 representa a aplicação dos modelos de CLV estudados. Neste caso, FC1, FC2, FC3, FC4 e FCn representam as transações efetuadas com um cliente, nos períodos 1, 2, 3, 4 e n. As setas apontadas para cima representam receitas; as setas apontadas para baixo, os custos. O ganho da transação é resultante entre as forças de pagamento e recebimento. Logo após, os ganhos são calculados em valor presente líquido. Este fluxo caracteriza uma simetria entre o recebimento das receitas e o pagamento dos custos.

É comum no contexto estudado as empresas comprarem e venderem seus produtos/serviços a prazo. Esta prática contribui para a assimetria entre os prazos de pagamento e recebimento.

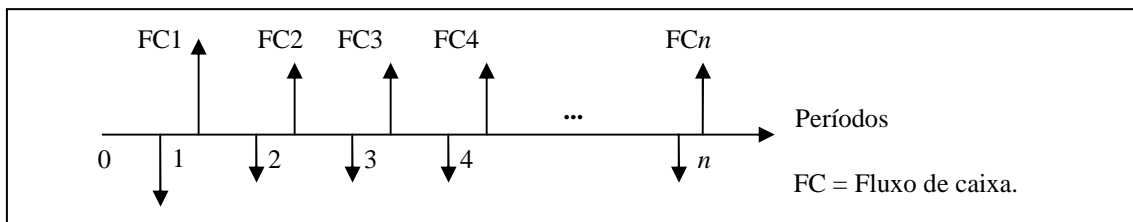


Figura 3. Fluxo de caixa com assimetria entre os prazos de pagamento e recebimento.

Na Figura 3, há diferenças entre os pagamentos dos custos e os recebimentos das receitas. Supondo que existam “altas taxas de juros”, esta assimetria causa impacto nos resultados do modelo de CLV. Para solucionar a assimetria nos prazos de pagamento e recebimento (venda a prazo) existente no contexto da empresa estudada, o modelo será:

$$CLV_j = \sum_{t=1}^T \left(\frac{q_j P}{(1+i)^{s_j}} \right) - \left(\frac{q_j C}{(1+i)^z} \right) \quad (19)$$

Assim:

CLV_j = o CLV do cliente j ;

j = índice do cliente;

t = índice do período;
 q_j = a quantidade comercializada pelo cliente j ;
 P = o preço unitário de venda do diesel, em reais;
 C = o preço unitário de compra do diesel, em reais;
 T = horizonte de previsão do CLV;
 i = custo do capital para a empresa;
 s_j = prazo concedido nas vendas para o cliente j ;
 z = prazo de pagamento dos fornecedores.

A solução para o problema é utilizar dois diferentes fatores de desconto: $(1+i)^{s_j}$ e $(1+i)^z$. O primeiro fator será utilizado para corrigir o valor das receitas ($q_j P$), e o segundo para os custos ($q_j C$). Esta solução se aproxima da utilizada no modelo de Kumar *et al.* (2008) (modelo 3, página 21). Kumar *et al.* (2008) também utilizam dois fatores de desconto, sendo um para a margem de contribuição do cliente e outro para os custos relacionados aos contatos (toques) de mercado. O modelo proposto se diferencia do modelo de Kumar *et al.* (2008), porque, no primeiro, a margem é dada após descontar os valores futuros da receita e despesa, ao passo que, no segundo, a margem é dada antes de se fazer tal desconto.

É importante ressaltar que, neste estudo, o termo “alta taxa de juros” é definido como as taxas praticadas atualmente no contexto brasileiro (10,25% a.a. meta Selic) (BACEN, 2010). Taxa de juros baixa seria a praticada no contexto americano, definida pelo Federal Reserve (Fed) como flutuante entre zero a 0,25% ao ano (ESTADÃO, 2010). Falar sobre assimetria entre pagamentos e recebimentos faz mais sentido se as taxas de juros forem altas (contexto brasileiro), pois o dinheiro tem valor no tempo, em outras palavras, considera que uma unidade monetária expressa em qualquer tipo de moeda tem diferentes valores (poder de compra da moeda), em diferentes datas (GITMAN, 2008). Por esse motivo, considerar a assimetria entre os recebimentos é mais importante no contexto brasileiro do que no contexto americano.

A abordagem adotada no modelo proposto é porque há a venda a prazo, a qual causa assimetria nos prazos de pagamento dos custos para os fornecedores e no recebimento da receita dos clientes. Na venda a prazo, a de se considerar o risco de não-recebimento desta receita, e quanto maior for o prazo de recebimento, maior é o risco, porque é dado mais prazo para que o cliente se torne insolvente (GITMAN, 2008).

Com o objetivo de antecipar clientes que no futuro serão inadimplentes, são encontrados na literatura os modelos de risco de crédito, os quais utilizam diferentes técnicas multivariadas baseadas em correlação, tais como: regressão logística, regressão discriminante e modelos de probabilidade linear (EIFERT, 2003; WILSON, SUMMERS e HOPE, 2000; ALTMAN, BAIDYIA e DIAS, 1979; BEAVER 1966; KANITZ, 1974; MINARDI, 1998 e SANVICENT). Nestes modelos, são utilizados como entrada dados extraídos da contabilidade (índice de liquidez, liquidez seca, entre outros) e, em alguns casos, são utilizados dados sobre a idade da empresa, o setor onde está inserida, entre outros.

Embora instituições financeiras exijam dados contábeis de seus clientes para análise de concessão de crédito, esta não é uma prática comum para empresas que atuam no comércio ou indústria na concessão de crédito comercial. Nessas situações, empresas que financiam clientes, através da concessão do crédito comercial (venda a prazo), possuem apenas dados cadastrais. Em casos em que o cliente tem uma relação de longa data com a empresa, ela também pode possuir dados históricos sobre o comportamento de pagamento do cliente.

Dados sobre o comportamento de pagamento passado e dados cadastrais (setor, idade da empresa cliente, entre outros) podem ser usados para prever o comportamento de pagamento futuro (WILSON, SUMMERS e HOPE, 2000). Neste intuito, esta dissertação irá mensurar o valor do cliente, incluindo a probabilidade de este tornar-se inadimplente, a partir da utilização de dados de comportamento de pagamento e dados cadastrais. Esta probabilidade é incluída na fórmula de CLV, através de uma variável μ_j . Assim, o modelo pode ser reescrito da seguinte forma:

$$CLV_j = \sum_{t=1}^T \left(\frac{\mu_j (q_j P)}{(1+i)^{s_j}} \right) - \left(\frac{q_j C}{(1+i)^z} \right) \quad (20)$$

Em que:

CLV_j = o CLV do cliente j ;

j = índice do cliente;

t = índice do período;

q_j = a quantidade comercializada no tempo t , pelo cliente j ;

μ_j = probabilidade de inadimplência do cliente j , no período t ;

P = o preço unitário de venda do diesel, em reais;

C = o preço unitário de compra do diesel, em reais;

T = horizonte de previsão do CLV;

i = custo do capital para a empresa;

s_j = prazo concedido nas vendas para o cliente j , no período t ;

z = prazo de pagamento dos fornecedores.

A probabilidade de inadimplência (μ_j) é um número qualquer entre zero e um, sendo zero inadimplência, e um adimplência. O μ_{jt} irá multiplicar a receita gerada pelo cliente j no período t . Caso ele possua um alto risco de inadimplência (μ_j próximo de zero), o índice punirá a receita, diminuindo o valor do CLV. Quanto maior o risco de inadimplência (μ_j próximo de 1), menor será o CLV do cliente.

A técnica multivariada, a qual será discutida no próximo tópico, que será utilizada para obter o índice μ_j será a regressão logística. O objetivo deste modelo é ajudar a empresa a ranquear seus clientes, podendo, assim, melhor definir seus clientes-alvo, escolhendo os que possuam boas margens de ganho futuro. A inclusão do risco de inadimplência no modelo de CLV facilita para a empresa decidir quais são as melhores escolhas. Uma vez decididos os clientes-alvo (maior CLV), ela pode melhor definir as políticas promocionais, tais como a adequação de descontos e prazos concedidos.

A empresa estudada concede percentuais de desconto sobre o preço de venda do diesel. O desconto varia com relação ao cliente. Os percentuais de desconto variam entre 3% e 9%. Os descontos são importantes de serem mensurados já que causam impacto no valor do CLV do cliente. Ao incluir o desconto no modelo, a fórmula será:

$$CLV_j = \sum_{t=1}^T \left(\frac{\mu_j (q_j P - d_j)}{(1+i)^{s_j}} \right) - \left(\frac{q_j C}{(1+i)^z} \right) \quad (21)$$

Assim:

CLV_j = o CLV do cliente j ;

j = índice do cliente;

t = índice do período;

q_j = a quantidade comercializada pelo cliente j ;

μ_j = probabilidade de inadimplência do cliente j ;
 P = o preço unitário de venda do diesel, em reais;
 C = o preço unitário de compra do diesel, em reais;
 T = horizonte de previsão do CLV;
 i = custo do capital para a empresa;
 s_j = prazo concedido nas vendas para o cliente j ;
 z = prazo de pagamento dos fornecedores;
 d_j = desconto do cliente j , no período t .

Igualmente o modelo pode ser reescrito da seguinte forma:

$$CLV_j = \sum_{t=1}^T q_j \left[\mu_j P \frac{1-d_j}{(1+i)^{s_j}} - \frac{C}{(1+i)^z} \right] \quad (22)$$

Ou seja:

CLV_j = o CLV do cliente j ;
 j = índice do cliente;
 t = índice do período;
 q_j = a quantidade comercializada pelo cliente j ;
 μ_j = probabilidade de inadimplência do cliente j ;
 P = o preço unitário de venda do diesel, em reais;
 C = o preço unitário de compra do diesel, em reais;
 T = horizonte de previsão do CLV;
 i = custo do capital para a empresa;
 s_j = prazo concedido nas vendas para o cliente j ;
 z = prazo de pagamento dos fornecedores;
 d_j = desconto do cliente j .

Os modelos de CLV diferem de acordo com o contexto e o tipo de empresa (JAIN e SINGH, 2002). Bons modelos são difíceis de encontrar por inúmeros motivos, sendo os principais: poucos deles possuem qualidade, a existência de modelos muito complexos, modelos incompletos (que não reproduzem inteiramente a realidade) entre outros (LITTLE,

2004). Embora o modelo proposto por esta pesquisa seja restrito e focado para o mercado varejista de combustíveis, seus principais pressupostos (risco de crédito, assimetria entre os prazos de pagamento e recebimento e desconto nas vendas), com alguns ajustes, podem ser aplicados a outros tipos de negócio nos quais existam as práticas de venda a prazo. Dentre os ajustes necessários, destaca-se a estimação de um novo modelo de risco de inadimplência (escolha das variáveis e critérios de avaliação e ajuste do modelo), verificação da necessidade de inclusão da variável de desconto, entre outros.

Existem pontos importantes para serem abordados em um modelo de CLV. Um desses pontos é a forma como o CLV processa os dados para prever o futuro. Diversos autores propõem modelos para mensurar o comportamento futuro dos clientes, os quais podem prever o que irá acontecer com a carteira de clientes nos próximos períodos. O modelo Pareto/NBD, de Schmittlein, Morrison e Colombo (1987), mensura a probabilidade de os clientes estarem ativos para a empresa nos próximos ciclos. Fader, Hardier e Jerath (2007) propõem outro modelo, baseado no Pareto/NBD, o RFM (recentidade, frequência e valor monetário), que é utilizado para mensurar o comportamento futuro dos clientes em um ambiente não-contratual e serve como entrada para os modelos de mensuração do CLV. É um modelo pouco utilizado devido à sua complexidade de aplicação. Visando solucionar este problema, Fader, Hardier e Lee (2005) propõem outro modelo, o BG/NBD, que é uma variação do modelo Pareto/NBD. Segundo os autores, aquele, comparado a este, é de simples aplicação, porém possui resultados preditivos piores.

Outra possibilidade é utilizar técnicas de *forecasting*, tais como técnicas multivariadas e de previsão de séries temporais (TUBINO, 2000). As técnicas de *forecasting* são importantes neste trabalho, pois representam uma alternativa para mensurar as compras futuras dos clientes.

Existem diversas possibilidades de ferramentas e métricas que podem ser utilizadas no modelo proposto. Porém para tornar este trabalho possível em um período de mestrado é necessário adotar algumas restrições. São elas: (1) o modelo irá medir o CLV dos clientes existentes na empresa (desconsiderando o crescimento ou a diminuição da base de cliente no futuro); (2) será mensurado o CLV para empresas que trabalham com ordem de frete (desconsiderando empresas que trabalham com compra à vista, convênios, cartão de crédito e cheque); (3) o modelo irá abordar apenas a venda de óleo diesel, sendo este o principal produto dos postos de rodovia; (4) o preço de venda (P) será constante e é aquele praticado atualmente pela empresa estudada; (5) o preço de compra (C) do diesel será constante e é uma média dos preços praticados pelas companhias (Petrobras, Shell, Esso e Ipiranga) que fornece

diesel para a empresa estudada; (6) a taxa de juros (i) será constante sendo a média do custo do capital da empresa; (7) foram desconsideradas as ações dos concorrentes; (8) no cálculo do risco, o modelo não irá considerar a aversão ao risco, e (9) o horizonte de cálculo do CLV será 1 ano (12 meses).

A seguir será discutida a técnica de regressão logística e como ela foi utilizada no modelo proposto nesta pesquisa.

2.4.1 Regressão logística

A regressão logística é uma técnica estatística multivariada. Ela tem como objetivo estimar a probabilidade associada à ocorrência de determinado acontecimento a partir de um conjunto de variáveis explicativas. Esta técnica é utilizada para situações em que a variável dependente possui natureza binária, podendo as variáveis independentes ser binária, dicotômica ou escalar (HAIR et al. 2005a; DIAS FILHO, 2003; MELLO e SLOMSKI, 2006).

Em comparação à análise discriminante e regressão linear, a regressão logística é mais robusta a violações de hipóteses (não exige normalidade multivariada e de igualdade de variância e covariância) (HAIR *et al.*; 2005a; BARTH, 2004). A não-exigência da normalidade torna a regressão logística mais adequada para os casos em que existam variáveis independentes não-métricas (adimplente/inadimplente, porte da empresa, entre outros). Para variáveis independentes não-métricas, é utilizada a técnica de *dummy variables*.

O modelo de regressão logística assume a seguinte expressão matemática (HAIR *et al.*; 2005):

$$\left(\frac{p}{1-p} \right) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots b_nx_n \quad (23)$$

Ou seja:

p = probabilidade de ocorrer adimplência;

$(1 - p)$ = probabilidade de ocorrer inadimplência;

x_n = variáveis independentes;

b_n = coeficientes atribuídos às variáveis independentes;

Alternadamente, a expressão pode ser reescrita como:

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + \dots + b_n x_n)}} \quad (24)$$

Este modelo tem como resposta um número maior ou igual a zero e igual ou menor que um. Ele prevê forçosamente uma probabilidade de adimplência entre zero e um, de acordo com a curva logística em formato de “S” (Figura 4). (BARTH, 2004).

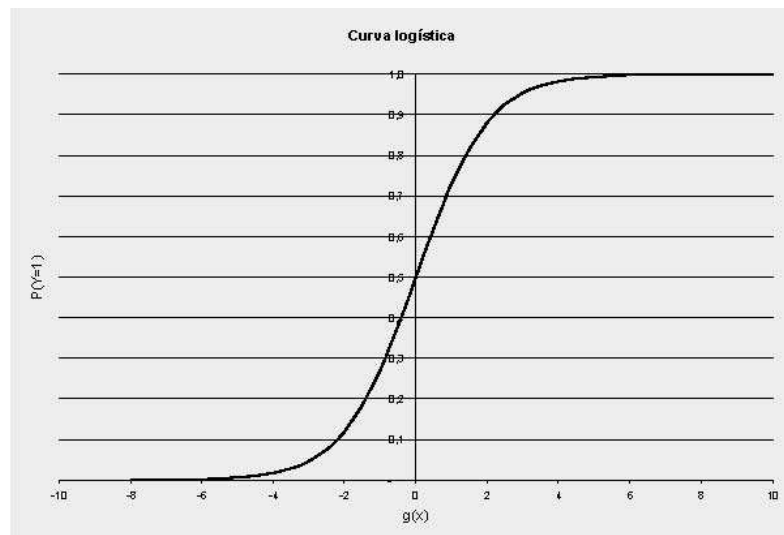


Figura 4. Curva logística.

Em níveis baixos da variável independente (y), a probabilidade tende a ser zero. Por outro lado, quando há um aumento na variável independente, a probabilidade tende a aumentar para cima da curva, mas há uma diminuição na inclinação da curva. Assim, a probabilidade poderá tender a um, porém nunca irá exceder este número.

Com relação à amostra, Hair *et al.* (2005) sugerem a utilização de no mínimo 5 ocorrências para cada variável dependente, sendo o recomendável 15. De acordo com os critérios de estimação utilizados no modelo (*forward*, *backward* e *stepwise*), esse número deve ser de 50 para 1.

O modelo *logit* necessita de mecanismos de avaliação dos resultados, que são estimados utilizando o método da máxima verossimilhança, sendo: *Cox-Snell R Square*, *Nagelkerke R Square*, *Homer e Lemeshow* e o teste *Wald*. (HOSMER E LEMESHOW, 2000; MELLO E SLOMSKI, 2006). O teste de *Cox-Snell R Square* é um teste semelhante ao *R*

Square utilizado na regressão linear. Ele não serve para indicar a proporção da variação da variável dependente em função das variáveis regressoras. Esse teste é baseado no *Likelihood Value* e possui uma escala que inicia em zero, mas não chega a 1 no limite superior. Por esse motivo, o teste de *Cox-Snell R Square* é utilizado para comparar modelos concorrentes. Inicialmente, entre dois ou mais modelos logísticos válidos, deve-se ter a preferência ao que apresentar o teste de *Cox-Snell R Square* mais elevado. O índice *Nagelkerke R Square* possui uma escala que vai de zero a um. O *Nagelkerke R Square* é um índice semelhante ao *Cox-Snell R Square*, porém é mais fácil de ser interpretado devido à existência de limites inferiores e superiores (HAIR *et al.* 2005a; MELLO e SLOMSKI, 2006).

O teste de *Hosmer e Lemeshow* é um indicador que corresponde ao teste do Qui-quadrado. Este teste separa o total de observações em dez classes e compara os resultados preditivos com os observados. Aqui existe um teste de hipóteses, no qual o objetivo é verificar a existência de diferenças estatisticamente significativas entre os resultados previstos e os dados observados. Diferentemente do normal, neste teste, o desejo é o de aceitar a hipótese nula. A interpretação do teste é que, se existem diferenças significativas, o modelo logístico não é capaz de produzir resultados satisfatórios (HOSMER e LEMESHOW, 1989; MELLO e SLOMSKI, 2006).

A estatística *Wald* tem o objetivo de atribuir um grau de significância estatística para cada coeficiente do modelo logístico. Esse teste é semelhante ao teste *t* e tem por objetivo verificar se cada parâmetro estimado é significativamente diferente de zero (HOSMER e LEMESHOW, 1989; MELLO e SLOMSKI, 2006).

A Figura 5 representa um modelo conceitual desta pesquisa. Os preços praticados, a taxa de juros e os descontos serão utilizados como *imput* para o modelo. Os históricos de compras dos clientes serão utilizados para prever as quantidades demandadas futuras. O risco de inadimplência será obtido através da regressão logística, utilizando-se dados de comportamento de pagamentos passados dos clientes (pagamento de faturas), e dados cadastrais da empresa (cidade, estado, entre outros).

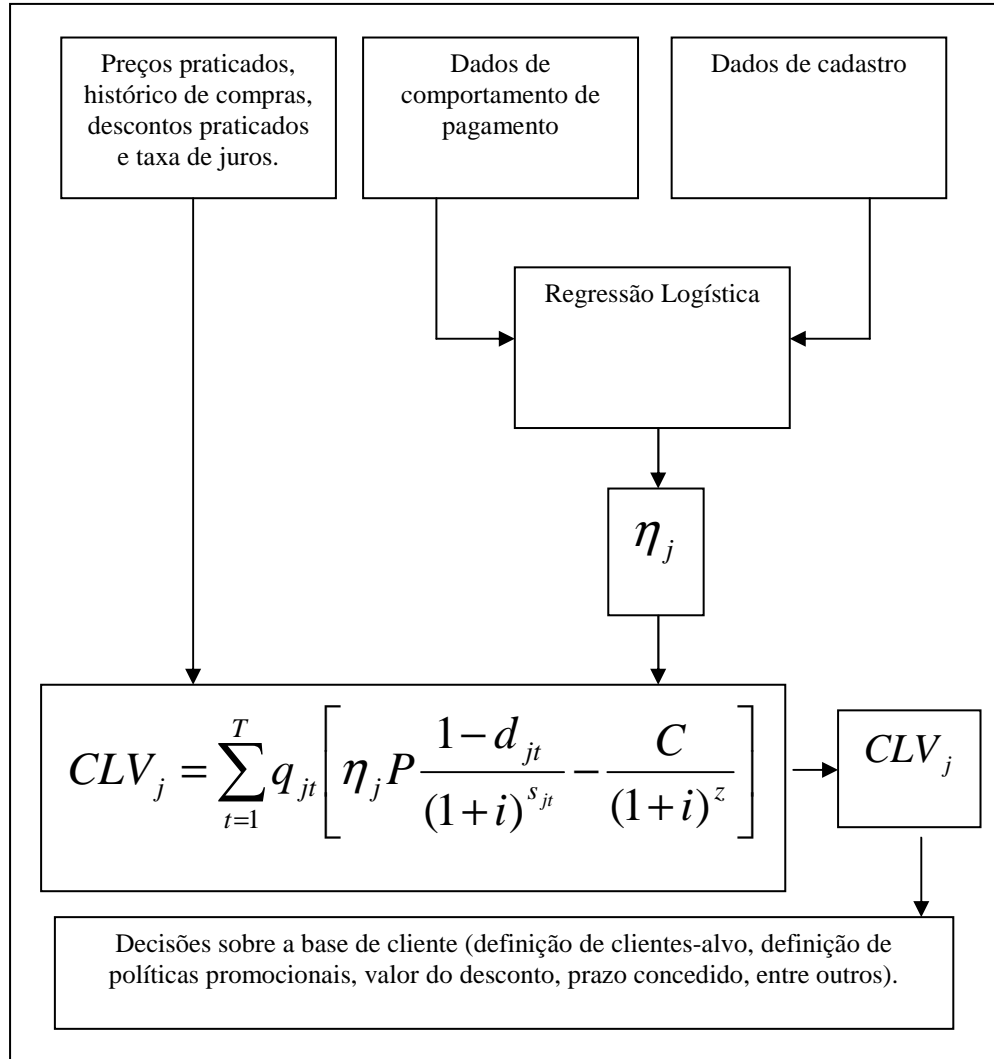


Figura 5. Modelo conceitual da pesquisa.

Entendendo a inclusão do risco de inadimplência como uma das principais contribuições do modelo, essa é a principal discussão nesta pesquisa. As outras variáveis, embora importantes, foram aqui suprimidas e trabalhadas de forma genérica. No capítulo 5 serão discutidas outras possibilidades para melhorar os resultados do modelo.

3. MÉTODO

O método da pesquisa é dividido em três fases, delimitadas de acordo com o objetivo proposto. As fases são: (1) fase exploratória; (2) fase de proposição do modelo e do método e (3) a fase de aplicação prática. A Figura 6 é uma representação das três fases da pesquisa, sua ordem e como estas se relacionam.

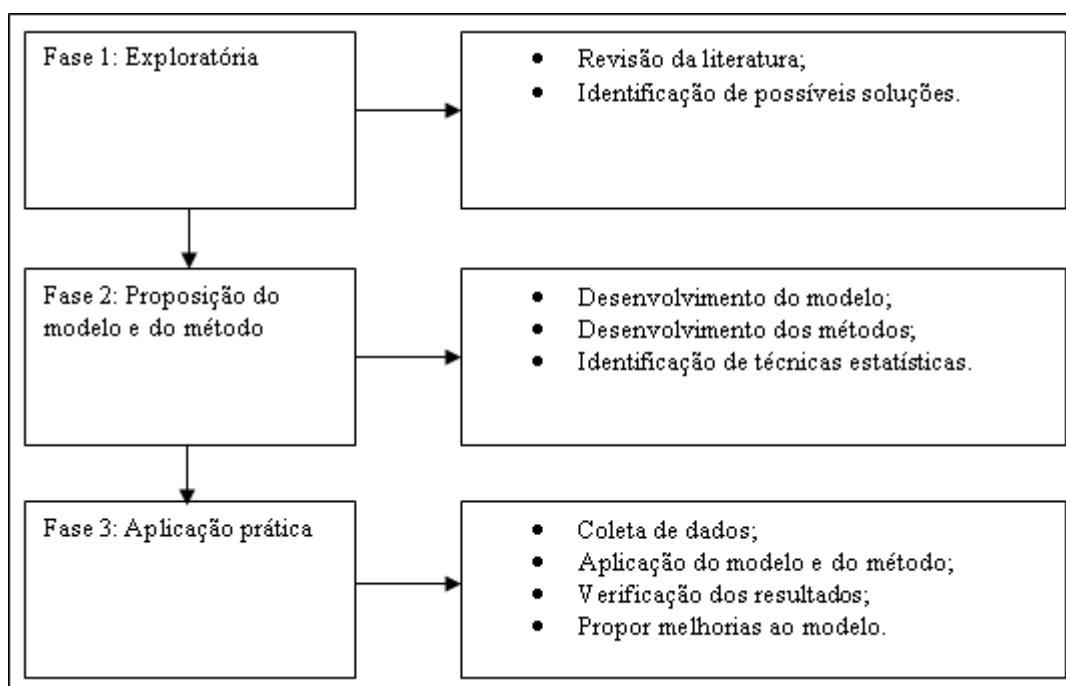


Figura 6. Método de pesquisa.

A primeira fase é caracterizada por não possuir procedimentos metódicos definidos. Aqui, o objetivo é uma ambientação com o assunto. Verificar se o problema já foi resolvido pela literatura e como esta aborda os conceitos-chave envolvidos na pesquisa. A partir desta fase, são desenvolvidas as bases teóricas para suportar o desenvolvimento da dissertação (MALHOTRA, 2006). Aqui foram pesquisadas principalmente duas áreas de conhecimento: *marketing* e finanças. As revisões da literatura associada a essas áreas permitiram identificar soluções parciais para o problema desta pesquisa. Dentre estas soluções parciais podem ser destacados principalmente o CLV da área do marketing e os modelos de previsão de inadimplência da área de finanças. O resultado da fase exploratória foi uma proposta de solução que integra uma relação de conceitos (CLV), técnica (Logit) e ferramentas (modelos de previsão de inadimplência) para solucionar o problema proposto, originando uma solução

não encontrada até então na literatura. Outra contribuição é a relação dos conceitos e técnicas aqui propostos e a forma como **esses** interagem para resolver o problema de pesquisa.

A segunda fase, denominada de proposição do modelo e do método, serviu para apontar, dentre as soluções levantadas, as possíveis alternativas levantadas na fase exploratória. **Essa** fase também ajudou a avaliar os pontos fortes e fracos de cada modelo e método, contribuindo para o amadurecimento da solução proposta. A revisão bibliográfica permitiu a construção do modelo matemático e também ajudou a levantar uma série de técnicas necessárias para a obtenção dos dados a fim de alimentar o modelo.

A terceira fase do método foi um estudo de campo, também chamado de estudo de caso, o qual se caracteriza por possuir um número reduzido de participantes na amostra, sendo, neste caso, apenas uma única amostra, chamada de estudo de caso único. **Esse** produz uma grande quantidade de dados, devido a dois motivos: (a) à profundidade com que as investigações são realizadas, e (b) por utilizar a pesquisa *in loco*. O estudo de caso é uma pesquisa realizada no contexto real do fenômeno estudado, tendo ainda como característica buscar dados em diversas fontes (YIN, 2001). Essa fase serviu principalmente como uma chamada de “volta à realidade”, permitindo confrontar a solução pretendida (o que se quer) com o que era possível de ser realizado devido às restrições na base de dados fornecidos pela empresa (o que se pode). O objetivo dessa etapa foi o de aperfeiçoar o modelo e o método proposto de forma que este possa ser validado no estudo de campo.

3.1. DEFINIÇÃO DA AMOSTRA

Aqui serão discutidos o setor e a empresa estudada, para introduzir o leitor no ambiente da pesquisa. Além disso, serão apresentados os tipos de dados coletados, seu total de observações e os procedimentos adotados na coleta.

3.1.1. O Setor

O setor escolhido para estudo de campo desta pesquisa são os postos de combustíveis de rodovia da região de Rio Grande e Pelotas. Essa escolha é orientada devido à conveniência do autor da pesquisa e de sua experiência no referido ramo.

Os postos de combustíveis de rodovia cumprem um importante papel na logística brasileira, principalmente porque o Brasil opta por um sistema de transporte para escoamento da produção agrícola baseado na utilização de rodovias. O modal rodoviário representa 61,2% do transporte nacional (FAEP, 2009), e é caracterizado por possuir um alto custo comparado aos demais modais. Esses custos são evidenciados na permanente necessidade de capital demandado para cobrir custos de viagens (pedágios, alimentação, combustíveis, mecânica, pneus, lubrificantes, filtros etc). Desta forma, os postos de rodovia, além de ofertar aos caminhões os produtos/serviços necessários para completarem a viagem, são encarregados também de suprir o capital necessário para cobrir os gastos de viagem. Em outras palavras, os postos de combustíveis de rodovia atuam hoje como instituições financeiras. Isso ocorre devido ao fato de o sistema de pagamento de considerável parte dos abastecimentos com óleo diesel se darem através do sistema de ordens de frete.

A ordem de frete é semelhante a uma nota promissória ou um cheque. E é emitida por uma transportadora em troca de serviços de frete efetuado por um caminhoneiro terceirizado. Ela é aceita como pagamento pelos postos de combustíveis de rodovia na medida em que os produtos e serviços são vendidos. Este sistema gera diversos problemas de caixa nos postos de combustível de rodovia, pois a maioria das ordens de frete possui valores superiores ao abastecimento. Em média, o valor da ordem de frete é de duas a três vezes o valor do abastecimento, sendo necessário que o posto efetue um troco (em cheque e/ou em dinheiro) superior ao valor dos combustíveis vendidos aos motoristas de caminhão. A Figura 7 mostra como funciona a relação de prestação de serviços com pagamento de ordem de frete.



Figura 7. Venda com ordem de frete.

Uma típica venda de combustível com ordem de frete é representada pelo seguinte caso:

- (+) Ordem de frete: R\$ 2.000,00
- (-) Abastecimento: R\$ 800,00
- (-) Troco em dinheiro: R\$ 200,00
- (-) Troco em cheque troco: R\$ 1.000,00

Neste exemplo, o caminhoneiro abastece R\$ 800,00 e entrega ao posto de combustível uma ordem de frete no valor de R\$ 2.000,00. Uma parcela do troco (R\$ 200,00) é em dinheiro, assim o caminhoneiro pode cobrir demais despesas de viagem (alimentação, estadia, entre outros). A segunda parte do troco é em cheque troco (cheque à vista). O cheque troco é um procedimento adotado pelos postos de rodovia por questões de segurança, isso diminui a quantidade de dinheiro nos caixas.

É comum que as ordens de frete emitidas tenham um prazo de recebimento, o qual pode variar de 3 a 60 dias, dependendo da carga e da negociação entre os postos e as transportadoras. Também é comum que os postos de rodovia deem descontos aos caminhoneiros. Os descontos variam de acordo com uma série de motivos, sendo: o cliente

(caminhoneiro/transportadora), o prazo de recebimento da ordem de frete, a quantidade abastecida, entre outros.

O sistema de ordens de frete causa uma série de problemas gerenciais nos postos de rodovia. Os problemas são desde o âmbito financeiro até problemas com o gerenciamento na carteira de clientes. Esses problemas podem ser mais bem entendidos se for analisada a empresa estudada.

3.1.2. A Empresa

A Abastecedora de Combustíveis Ongaratto Ltda. (Postos Ongaratto) é o principal ramo de negócio do Grupo Ongaratto, que iniciou suas atividades em 1986, no estado do Rio Grande do Sul. Os Postos Ongaratto atuam no ramo varejista de combustíveis, lubrificantes e prestação de serviços para veículos leves e pesados. Atualmente, a empresa possui uma rede composta por 16 postos, sendo 5 de postos de rodovia (veículos pesados) e 11 postos de cidade (veículos leves). Quatro dos cinco postos de rodovia estão situados na BR 392, próximos ao Porto de Rio Grande, uma das principais vias de escoamento da safra de grãos brasileira.

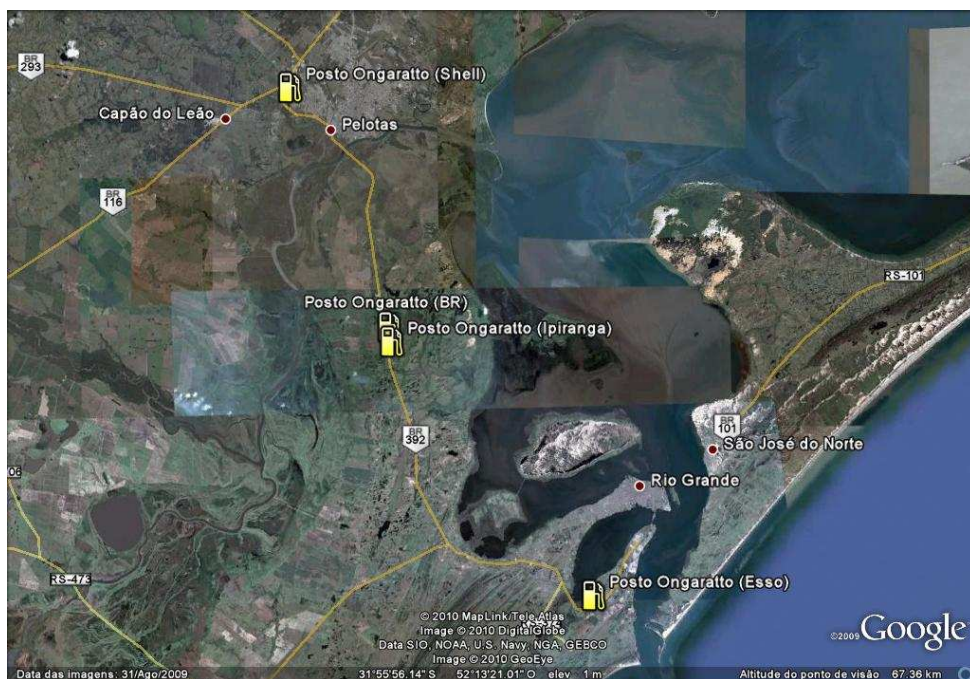


Figura 8. Localização dos postos de combustível.
Fonte: Google Earth.

O estudo de campo desta pesquisa foi focado apenas nos postos de rodovia (5 unidades) e especificamente para clientes que são transportadoras e que trabalham com a emissão de ordens de frete. O motivo desta delimitação é baseado na dificuldade encontrada pela empresa estudada em ranquear seus clientes (transportadoras) e em definir níveis adequados de desconto e prazo de recebimento. Essa dificuldade ocorre devido a uma série de fatores, sendo: a utilização de métodos empíricos para o ranqueamento, a complexidade da operação de venda com a ordem de frete e a existência de expressiva inadimplência no pagamento das ordens de frete.

Em 2008, os Postos Ongaratto emitiram 8.759 faturas com um total de R\$ 88.882.593,76 apenas com ordens de frete. Deste total, 2.291 faturas ou R\$ 26.895.604,42 foram pagos com atrasos iguais ou superiores a 7 dias. Se fossem somados todos os dias de atraso de cada fatura classificada como inadimplente, o total seria 57.345 dias. Considerando o custo do capital para a empresa, o juro total que deveria ser cobrado dos inadimplentes seria R\$ 529.882,50 (sem considerar multas e demais custos) apenas no ano de 2008. Deste total foram resgatados apenas R\$ 21.856,59.

Atualmente, o total de empresas na carteira de clientes que trabalham com ordem de frete é 229 transportadoras. Do total, apenas 7 clientes detêm 47, 5% do faturamento com ordem de frete.

Cliente	Faturamento (R\$) 2007 a 2009	% do Faturamento
49 e 1568	18.781.462,36	12,97
605 e 844	12.375.673,67	8,54
144	10.545.887,75	7,28
245	8.535.697,71	5,89
594	7.831.258,55	5,41
145	5.438.163,73	3,75
589	5.286093,62	3,65

Tabela 4. Faturamento por cliente de dezembro 2007 a agosto 2009.

Nesta pesquisa, os clientes são tratados apenas por códigos numéricos. Apenas dois clientes (“49 e 1568” e “605 e 844”) possuem dois códigos de referência na base de dados. Isso aconteceu devido a mudanças efetuadas no cadastro. Na Tabela 4, estão ranqueados os sete clientes com maior faturamento.

Para a empresa estudada é importante ranquear seus clientes (transportadoras), pois isso ajuda na definição de promoções, prazos de pagamento e descontos. Na maioria das situações, na operação de venda envolvendo ordem de frete é concedido ao cliente um prazo de pagamento de 15 dias com desconto de 5% no preço do combustível. Neste exemplo, a venda acontece da seguinte maneira: a transportadora e posto de combustível chegam a um acordo de parceria. Este acordo pode resultar em um contrato por escrito ou verbal de compra/venda. Neste acordo, o posto se compromete a abastecer os caminhões (terceirizados), aceitando como pagamento as ordens de frete emitidas pela transportadora, para a qual é acertado também um prazo de pagamento.

O posto de combustível, quando acontece o abastecimento, geralmente estipula uma norma de que o caminhoneiro tenha de abastecer 40% do valor total da ordem de frete. Caso ele tenha uma ordem de frete de R\$ 1.000,00, R\$ 400,00 será gasto em abastecimento e R\$ 600,00 reais serão dados de troco. É concedido um desconto de 5% no abastecimento (R\$ 20,00), sendo este entregue ao caminhoneiro.

Após o abastecimento, o posto envia as ordens de frete a um escritório central, onde estas são agrupadas com outras ordens da mesma transportadora que chegam de outras filiais (postos) da empresa. As ordens são agrupadas e faturadas. A fatura é enviada à transportadora e o pagamento é agendado conforme prazo acordado anteriormente. O prazo mais comum é o de 15 dias. Porém este prazo pode variar de três até 60 dias. Neste trabalho, o pagamento com inadimplência refere-se ao atraso no pagamento de faturas.

3.1.3. Período de Análise e Número de Observações

Esta pesquisa utiliza dados das seguintes fontes:

Fonte	Descrição da fonte	Tipo de dado coletado
Relatório de faturamento	Relatório diário com dados de todas as faturas emitidas pelo setor de faturamento da empresa	Dados históricos de compras dos clientes, preços de venda por cliente, prazo de pagamento por cliente e percentual de desconto por cliente.
Relatório de recebimento	Relatório diário contendo as faturas pagas e não pagas pelos clientes	Faturas pagas em dia ou com atraso. Faturas com falta de pagamento
Relatório de preços	Relatório com os preços de vendas dos produtos discriminado por filial	Preços de venda do diesel discriminado por filial.
Cadastro de clientes	Relatório com informações sobre os clientes	Dados cadastrais dos clientes.
Notas fiscais de compras		Valor de custo do combustível para cada filial.
Especialistas	Pessoas envolvidas com a empresa estudada.	Cientes que ofereceram garantias, tipo de carga transportado pelos clientes, prazo médio de atraso.

Tabela 5. Coleta de dados.

Dentre as fontes citadas na Tabela 5, as principais foram o Relatório de faturamento, emitido pelo setor de faturamento da empresa, e o Relatório de recebimentos, emitido pelo setor financeiro. A junção dos dados destes dois relatórios permitiu criar uma base de dados, contendo todas as faturas emitidas entre dezembro de 2007 e setembro de 2009. Além disso, é possível identificar quais faturas foram pagas no prazo, quais foram pagas com atraso e quais não foram pagas.

O cadastro de clientes continha diversas informações, tais como a localização geográfica (cidade e estado), CNPJ, IE, telefone, endereço, placas da frota, entre outros. Destas informações, foram utilizados apenas os dados relacionados a cidade e estado. Os dados relacionados a frota, quantidade de caminhões e placas não foram utilizados devido à quantidade de dados faltantes existentes no cadastro.

Os especialistas contribuíram para completar dados faltantes relacionados aos prazos de pagamento, o tipo de carga com as quais as transportadoras trabalham (grãos, container entre outros) e também contribuíram com dados relacionados ao faturamento. O relatório de preços e as notas fiscais de compra foram utilizados para identificar o valor de compra e de venda do diesel, sendo estes duas importantes variáveis para o cálculo do CLV.

Estes dados foram agrupados em uma tabela, utilizando o Microsoft Office Excel 2007. Após a reunião dos dados, estes passaram por procedimentos de limpeza e organização, que serão descritos no próximo tópico.

3.2. COLETA E TRATAMENTO DOS DADOS

A base de dados desta pesquisa possui informações de pagamento de faturas de dezembro de 2007 a setembro de 2009. Em dezembro de 2007, a empresa estudada adquiriu um novo sistema operacional, restringindo a coleta a esta data. Como a coleta foi realizada em setembro de 2009, utilizou-se o maior horizonte de dados possível.

Os dados utilizados para mensurar o risco de inadimplência dos clientes foram coletados principalmente dos relatórios de faturamento e recebimento da empresa. Esses relatórios foram emitidos em Microsoft Office Excel 2007, sendo agrupados manualmente em apenas uma tabela. Essa tabela contém todas as faturas emitidas pela empresa e dentre os seus principais dados estão: o código de controle da fatura, a data de emissão, a data de vencimento, a data de pagamento, o tipo de pagamento utilizado pelo cliente (depósito em conta, boleto bancário, cheque, pagamento em dinheiro), o valor da fatura, descontos concedidos no pagamento, juros cobrados, número de dias em atraso, código do cliente, nome do cliente e responsável pela emissão da fatura.

Inicialmente, a tabela em Excel compreendia 42.384 linhas por 16 colunas. O relatório estava disposto de forma que a cada 2 linhas havia um registro de pagamento de fatura. Essa tabela continha dados de aproximadamente 21.692 faturas. Como o foco do trabalho é apenas para empresas que trabalham com ordem de frete, foram inicialmente realizadas duas mudanças na base, sendo: (a) identificação e exclusão de todas as faturas que não estão relacionadas com ordem de frete e (b) a base foi reorganizada de forma a conter em cada linha apenas uma fatura. Para identificar as faturas que não estavam relacionadas a ordens de frete era necessário inicialmente organizar toda a base. Essa reorganização foi realizada manualmente, de forma que cada linha contivesse dados de apenas uma fatura. Após esse trabalho, foram identificadas, através do recurso de autofiltro, todas as faturas de clientes que não trabalham com ordem de frete, sendo estas excluídas. O próximo passo foi substituir o nome de todos os clientes por códigos. Essas medidas reduziram a base de dados de 42.384 linhas e 21.692 faturas para 14.259 linhas e faturas, o valor total de faturamento da tabela ficou em R\$ 144.836.649,10 apenas com ordem de frete.

Com a nova base de dados, o próximo passo foi classificar quais faturas foram pagas em dia e quais eram inadimplentes. Para isso, foi calculada a diferença em dias da data de vencimento e a data de pagamento de cada uma delas. O resultado deste cálculo assume um valor para cada uma das 14.259 faturas, podendo ser (a) zero nos casos em que não há a

ocorrência de atrasos, (b) um número positivo caso exista atraso no seu pagamento e (c) um número negativo indicando que o cliente adiantou o pagamento à empresa. A partir desta variável foi constatado que aproximadamente 12,2% (1.744) dos pagamentos efetuados possuíam algum tipo de antecipação. Através de um estudo mais aprofundado, verificou-se que do total de pagamentos antecipados, 97% das ocorrências eram de clientes que pagam faturas utilizando o cheque pré-datado. Para efeito de controle interno da empresa estudada, quando um cliente paga uma fatura utilizando um cheque pré-datado, sua fatura é quitada no sistema e o seu recebimento é registrado na data de entrega do cheque. O registro que fica pendente no sistema é o cheque até a data de sua compensação. Assim, fica registrado no relatório como se a fatura estivesse com o pagamento antecipado, quando, na verdade, isso não ocorreu. Para resolver esta distorção na base, buscou-se primeiramente fazer um cruzamento dos relatórios de faturamento e de cheques, porém isso não foi possível devido a restrições nas bases de dados da empresa estudada. Outra possibilidade era excluir estes clientes da base de dados, porém eles possuem considerável número de faturas (1.268).

Dentre os clientes que apresentavam distorção nas datas de pagamento com cheque pré-datado, estavam: 7, 13, 128, 145, 156, 397, 589, 598 e 605 e 844. Para corrigir essas deformações na base de dados, foram entrevistados funcionários que trabalham no setor de crédito e cobrança e no setor financeiro da empresa com o objetivo de buscar um padrão de comportamento de pagamento para estes clientes. Assim foram identificados os clientes adimplentes e inadimplentes e também um prazo médio de recebimento (Tabela 6).

Cliente	Comportamento médio de pagamento	Dias de atraso médio
7	Assíduo/adimplente	0
13	Assíduo/adimplente	0
128	Atraso/Inadimplente	30
145	Atraso/Inadimplente	15
156	Atraso/Inadimplente	20
397	Atraso/Inadimplente	20
589	Atraso/Inadimplente	20
598	Atraso/Inadimplente	25
605e844	Atraso/Inadimplente	20

Tabela 6. Alterações realizadas na base de dados.

O prazo de atraso médio em dias foi aplicado a todos os pagamentos dos clientes que foram pagos em cheque pré-datado. O cliente 589 possuía parte dos pagamentos efetuados via

depósito bancário e também com cheque pré-datado. Nesse caso o prazo médio foi aplicado apenas aos pagamentos realizados através de cheque.

Finalizada as correções na base de dados, partiu-se para a classificação das faturas em adimplência e inadimplência. Neste trabalho foi definida a inadimplência como qualquer atraso no pagamento igual ou superior a 7 dias. Essa definição aproxima-se das definições utilizadas por Wilson, Summers e Hope (2000), que eram atrasos no pagamento iguais ou superiores a 15 dias e atrasos nos pagamentos iguais ou superiores a 30 dias. A utilização de uma definição de inadimplência com apenas 7 dias é justificada devido ao interesse da empresa estudada nos resultados. Para a empresa, uma definição de inadimplência mais sensível (apenas 7 dias) é justificável por causa da ordem de frete e o necessário troco (ver página 51) que causam impacto relativamente alto no saldo de caixa da empresa. Por esse motivo, pequenos atrasos no pagamento podem acarretar altos custos financeiros. Inicialmente, a ideia da empresa era a de classificar qualquer atraso como inadimplência. Porém isso não foi possível devido a fragilidades nos dados, pois o setor financeiro da empresa estudada concilia os pagamentos no dia posterior à data de pagamento efetuado pelo cliente. Assim podem acontecer distorções nas datas de vencimento e pagamento de 1 até 3 dias. Em um caso extremo, um pagamento efetuado por um cliente na sexta será liquidado junto à fatura apenas na segunda-feira no sistema da empresa. Dadas essas limitações no processo de liquidação de faturas da empresa, adotou-se um prazo de 7 dias de forma a ignorar falsos atrasos na base de dados. Foram testadas nesta pesquisa outras definições de inadimplência, sendo: (a) atrasos iguais ou superiores a 15 dias e (b) atrasos iguais ou superiores a 30 dias, porém os resultados desses testes não foram incluídos nesta pesquisa e são mais bem comentados no Capítulo 5.

Depois de adotado o critério de classificação de faturas inadimplentes (faturas pagas com 7 ou mais dias de atraso), a classificação foi realizada utilizando a função “SE” do Microsoft Office Excel 2007. Essa classificação foi realizada através da técnica de variáveis dummy. Aqui as faturas adimplentes (<7 dias) eram classificadas como “0” (zero) e as faturas inadimplentes (=7 ou >7 dias) eram classificadas como “1” (um).

Também foram adicionados à tabela dados do cadastro de clientes. Esses dados compreendiam a localização do cliente (cidade, estado) e se ele estava na condição de fiador e o tipo de carga mais transportado (grãos, contêiner, entre outros). Esses dados também foram acomodados com variáveis dummy.

A base de dados tinha 14.259 linhas por 29 colunas, e foram excluídas as ocorrências (faturas) com dados faltantes (643). Agora a base de dados tem um total de 13.616 linhas por

29 colunas. Esses dados foram utilizados principalmente para mensurar os coeficientes para o modelo de risco de inadimplência que integra o modelo de CLV. Esses procedimentos serão mais bem explicados no próximo capítulo.

4. ANÁLISE DOS DADOS

Neste capítulo serão abordados os procedimentos adotados para a estimação do modelo e análise dos resultados. Inicialmente serão analisados os resultados do modelo de inadimplência, que é integrado ao modelo de CLV, cujos resultados serão discutidos depois.

4.1. MODELO DE INADIMPLÊNCIA

Aqui serão descritos os procedimentos adotados para a construção do modelo de inadimplência. Serão discutidas as variáveis utilizadas, a técnica de estimação e os processos de calibração e projeção adotados.

4.1.1. Variáveis Explicativas ou Regressoras

Após organizar e limpar a base de dados, foram utilizadas as seguintes variáveis explicativas para construir o modelo de inadimplência:

Variável	Relação com a inadimplência
Pag1	Ocorrência de inadimplência no pagamento da fatura do período anterior (t-1).
Pag2	Ocorrência de inadimplência no pagamento da fatura do período anterior (t-2).
Pag3	Ocorrência de inadimplência no pagamento da fatura do período anterior (t-3).
Pelotas	Suposição de que clientes da mesma praça da empresa estudada tendem a ser menos inadimplentes.
Rio Grande	Suposição de que clientes da mesma praça da empresa estudada tendem a ser menos inadimplentes.
RS	Suposição de que clientes de outros estados tendem a ser mais inadimplentes.
Grãos	Empresas que trabalham com os grãos tendem a ser mais inadimplentes devido a flutuações no mercado.
Contêiner	Empresas que trabalham com contêiner tendem a ser mais inadimplentes.
Garantia Exigida	Para concessão do crédito comercial, a empresa estudada exige de alguns clientes uma garantia real. Aqui a relação é a de que os clientes em que a empresa percebe risco são mais inadimplentes que outros.
Vlr Recebido	Quanto maior o valor da fatura, maior a inadimplência.
Dias Atraso	Empresas com histórico de atraso tendem a atrasar mais no futuro.

Tabela 7. Variáveis predictoras testadas.

A maioria dos modelos de inadimplência utiliza (a) dados coletados de análises de balanço (BEAVER, 1966; ALTMAN, 1968; ALTMAN, BIDYA E DIAS, 1979; SANVICENT E MINARDI, 1998) e/ou e (b) dados sobre perfil do cliente (RIBEIRO, 2008). Poucos modelos utilizam dados sobre o comportamento de pagamento (WILSON, SUMMERS E HOPE, 2000). Esses dados são de difícil acesso devido a questões de sigilo nas empresas. Outra questão importante é que os modelos de inadimplência disponíveis na literatura são, em sua maioria, direcionados para instituições financeiras. Nesses casos, o modelo ajuda a empresa a decidir em conceder ou não crédito a um cliente, identificando futuras insolvências que impossibilitem o pagamento do empréstimo.

Neste trabalho, o modelo de inadimplência desenvolvido tem o objetivo de identificar maus pagadores. A insolvência, embora importante, é um resultado extremo de problemas financeiros. É provável que atrasos nos pagamentos de compromissos sejam indícios de futuras insolvências (GITMAN, 2008).

Nesta pesquisa foram utilizados dados que normalmente são encontrados nos sistemas de empresas que trabalham vendendo produtos/serviços a prazo (contexto de crédito comercial). São informações básicas sobre o perfil dos clientes (endereço, nome, condições de pagamento, entre outros) e um histórico de registro de pagamentos efetuados. As variáveis

testadas no modelo foram escolhidas utilizando a percepção do autor e sua crença na relevância da variável para o estudo.

Variáveis obtidas com a análise de balanço, embora sejam normalmente utilizadas em modelos de inadimplência, foram descartadas por três motivos: (a) muitos dos clientes da empresa estudada (transportadoras) são pequenas prestadoras de serviço (apenas uma sala e telefone) e não possuem sistemas estruturados que permitam um melhor acompanhamento dos dados de balanço, (b) informações contábeis raramente são solicitadas no mercado mencionado e (c) estas informações não são normalmente solicitadas dos clientes para empréstimos no contexto de crédito comercial.

4.1.2. Calibração e projeção

Para estimação do modelo, foi utilizada a técnica de regressão logística. Inicialmente foi estimado um modelo utilizando apenas 50% (6808) da base de dados. Isso foi possível devido à grande quantidade de dados disponíveis na amostra (13.616). O método utilizado foi o *enter*, o qual estima um modelo com a inclusão simultânea de todas as variáveis, independente de sua capacidade de explicação com a variável dependente. Esse modelo serviu basicamente para verificar a significância estatística das variáveis independentes em discriminar a inadimplência. Os resultados estão na Tabela 8 e Tabela 9, sendo:

Cox & Snell R Square	0,315
Nagelkerke R Square	0,458

Tabela 8. Resultados da estimação.

O teste de *Cox & Snell* é semelhante ao R quadrado da regressão linear. Isso indica que 31,5% das variações ocorridas no log da razão de chances são explicadas pelo conjunto das variáveis regressoras. Como o teste de *Cox & Snell* não possui limite superior, ele é mais indicado para comparação de modelos concorrentes. O teste Nagelkerk, semelhante ao teste de *Cox & Snell*, indica que o modelo é capaz de explicar 45,8% das variações registradas na variável dependente (HOSMER E LEMESHOW, 1989; MELLO E SLOMSKI, 2006).

Variáveis	Coefficiente de regressão	Significância estatística
Pag1	2,171	0,00
Pag2	1,422	0,00
Pag3	1,349	0,00
Pelotas	0,116	0,269
Rio Grande	0,430	0,00
RS	-0,084	0,692
Contêiner	-0,036	0,808
Grãos	0,289	0,029
Polietileno	-0,376	0,256
Vlr Recebido	-0,00001	0,00
Garantia Exigida	0,712	0,00
Constante	-2,321	0,00

Tabela 9. Resultados da estimação.

Na Tabela 9, as variáveis “Pelotas”, “RS”, “Contêiner” e “Polietileno” não se mostraram significativas estatisticamente para explicar a variável dependente, sendo significativo o restante das variáveis. O teste de *Hosmer e Lemeshow* indicou a existência de falta de ajuste no modelo (sig. = 0,02). Esse resultado reflete a existência de variáveis preditoras não-significativas no modelo.

O próximo passo foi estimar o modelo, empregando o método *forward LR*. Esse método de estimação inclui variáveis no modelo de acordo com sua capacidade de classificação com a variável dependente. O método *forward LR* inicia o processo de estimação com a variável que possui o maior poder discriminante. As variáveis vão sendo incluídas sucessivamente, podendo uma ou outra ser eliminada no caso de a interação de duas variáveis agregar maior poder discriminante (HAIR *et al*, 2005a). Apesar de este método ser amplamente utilizado nos processos de estimação com a técnica de regressão logística (EIFERT, 2003; MELLO E SLOMSKI; 2006; RIBEIRO, 2008), nesta pesquisa, foi também testado o método *backward LR*, que estima o modelo utilizando um processo contrário ao *forward LR*. Esses resultados serão mais bem comentados no Capítulo 5. O resultado do modelo está na Tabela 10 e Tabela 11.

Cox & Snell R Square	0,345
Nagelkerke R Square	0,508

Tabela 10. Resultados da estimação.

Conforme se vê na Tabela 10, 34,5% das variações ocorridas no log da razão de chances são explicadas pelo conjunto das variáveis regressoras. O teste Nagelkerk indica que o modelo é capaz de explicar 50, 8% das variações registradas na variável dependente (HOSMER E LEMESHOW, 1989; MELLO E SLOMSKI, 2006).

Variáveis	Coefficiente de regressão	Significância estatística
Pag1	2,173	0,000
Pag2	1,426	0,000
Pag3	1,351	0,000
Rio Grande	0,389	0,000
Grãos	0,369	0,000
Vlr Recebido	-0,00001	0,000
Garantia Exigida	0,714	0,000
Constante	-2,432	0,000

Tabela 11. Resultados da estimação.

Sendo essas as variáveis que compõem o modelo, o próximo passo foi verificar a existência de multicolinearidade nas variáveis regressoras. Ela ocorre quando há a existência de correlações altas ($r = 0,8$) entre as variáveis explicativas. A existência de multicolinearidade pode indicar que as variáveis estão medindo a mesma coisa. Isso causa deformações nos resultados dos modelos em termos de explicação e estimação (HAIR et al., 2005a; DANCEY E REIDY, 2006).

Uma forma de observar a sua existência é utilizando a matriz de correlação bivariada (r de Pearson). O propósito de uma análise de correlação é saber se duas variáveis estão relacionadas e isso ocorre quando elas covariam. A covariação é quando uma variável muda coerentemente e sistematicamente de acordo com outra variável. O coeficiente r de Pearson mede a correlação entre duas variáveis. O r de Pearson é um coeficiente que varia de -1 a +1. Coeficientes próximos a zero indicam pouca ou nenhuma correlação, ao passo que coeficientes próximos de +1 ou de -1 demonstram uma correlação forte. Correlações positivas indicam que as duas variáveis se comportam de maneira semelhante. Correlações negativas indicam que as variáveis se comportam de maneira inversamente proporcional, ou seja,

quando uma variável aumenta, a outra diminui. O coeficiente também é acompanhado por um teste de significância estatística. O teste de significância busca testar se existe uma relação linear entre as duas variáveis e se esta correlação aconteceu devido ao erro amostral. Normalmente coeficientes abaixo de 0,05 são considerados como estatisticamente significativos (HAIR et al. 2005b; DANCEY E REIDY, 2008).

O coeficiente de correlação r de Pearson é classificado de acordo com sua intensidade. Na Tabela 34 é descrita a classificação da intensidade do relacionamento conforme Dancey e Reidy (2008), sendo:

Classificação do relacionamento	Coficiente r de Pearson	
	Perfeito	1
Forte	0,9	-0,9
	0,8	-0,8
	0,7	-0,7
	0,6	-0,6
Moderado	0,5	-0,5
	0,4	-0,4
	0,3	-0,3
	0,2	-0,2
Fraco	0,1	-0,1
	0	0
Zero	0	0

Tabela 12. Classificação do coeficiente de correlação.

Fonte: Adaptado de Dancey e Reidy (2008).

O coeficiente de correlação foi calculado utilizando o SPSS 17. No Quadro 1, estão os resultados do teste, o qual demonstrou que não existem correlações acima de 0,7 entre as variáveis independentes da amostra. Com exceção das variáveis “Pag1”, “Pag2” e “Pag3”, todas as outras variáveis possuem um baixo grau de correlação, segundo critério utilizado por Hair *et al.* (2005b). Esses resultados indicam inicialmente que os dados não serão afetados por problemas de multicolinearidade.

		Pag1	Pag2	Pag3	Rio Grande	Grãos	Vlr Recebido
Pag1	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	1					
Pag2	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	0,683 0,00	1				
Pag3	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	0,612 0,00	0,682 0,00	1			
Rio Grande	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	-0,083 0,00	-0,084 0,00	-0,083 0,00	1		
Grãos	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	-0,122 0,00	-0,122 0,00	-0,122 0,00	0,015 0,08	1	
Vlr Recebido	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	0,085 0,00	0,085 0,00	0,084 0,00	0,227 0,00	-0,010 0,26	1
Garantia Exigida	<i>Pearson Correlation</i> <i>Sig. (2-tailed)</i>	0,229 0,00	0,232 0,00	0,233 0,00	-0,163 0,00	-0,230 0,00	0,105 0,00

Quadro 1. Matriz de correlação.

Embora a multicolinearidade seja mais comum na existência de níveis altos de correlações, ela também pode ocorrer em correlações fracas (DANCEY E REIDY, 2008; HAIR *et al.* 2005a). Outra forma de verificar a sua existência é através do teste de Fator de Inflação de Variância (FIV). O FIV mede o quanto a variância das variáveis independentes é afetada por problemas de multicolinearidade. No teste, a interpretação dos resultados é que valores baixos (de 0 a 5) indicam pouca associação entre as variáveis, mas geralmente não o suficiente para indicar algum tipo de problema. Na literatura, um valor considerado limite para o FIV é 5, 0, valores acima disso indicam problemas com multicolinearidade. Na Tabela 13 estão os resultados do teste:

Variável	FIV
Pag1	2,219
Pag2	2,560
Pag3	2,185
Rio Grande	1,118
Grãos	1,050
Garantia Exigida	1,090
Fiador	1,141

Tabela 13. Resultado do teste FIV.

As variáveis “Pag1”, “Pag2” e “Pag3” destacam-se novamente como possuindo algum indício de multicolinearidade, porém os índices de FIV estão abaixo do máximo aceitado (5,0) segundo critérios descritos por Hair *et al* (2005b). Desta forma, o modelo de previsão de inadimplência é composto pelo conjunto de sete variáveis vistas na Tabela 8 e Tabela 9, sendo:

- Pag1: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-1);
- Pag2: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-2);
- Pag3: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-3);
- Rio Grande: Clientes localizados na praça da empresa Rio Grande;
- Grãos: Clientes que trabalham com cargas a granel (soja, arroz entre outros);
- Vlr Recebido: Valor da fatura paga pelo cliente;
- Garantia Exigida: Clientes de quem a empresa exigiu algum tipo de garantia para efetuar a venda a prazo.

A função matemática que correspondente ao modelo de previsão de inadimplência é a Equação (25), sendo:

$$\left(\frac{p}{1-p} \right) = -2,432 + 2,173Pag1 + 1,426Pag2 + 1,351Pag3 + 0,389RioGrande... \quad (25)$$

$$... + 0,369Graos + 0,0001Vlr Recebido + 0,714GarantiaExigida$$

Assim:

p = probabilidade de ocorrer inadimplência;

$(1 - p)$ = probabilidade de ocorrer adimplência;

Pag1: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-1);

Pag2: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-2);

Pag3: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-3);

Rio Grande: Clientes localizados na praça da empresa Rio Grande;

Grãos: Clientes que trabalham com cargas a granel;

Vlr Recebido: Valor da fatura paga pelo cliente;

Garantia Exigida: Clientes de quem a empresa exigiu algum tipo de garantia para efetuar a venda a prazo.

Da mesma forma, a expressão da função risco de inadimplência pode ser reescrita da seguinte forma:

$$p = \frac{1}{1 + e^{-(2,432 + 2,173Pag1 + 1,426Pag2 + 1,351Pag3 + 0,389RioGrande...)}} \quad (26)$$

$$\dots + 0,369Graos - 0,00001VlrRecebido + 0,714GarantiaExigida)$$

Onde:

p = probabilidade de ocorrer inadimplência;

Pag1: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-1);

Pag2: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-2);

Pag3: Ocorrência de inadimplência no pagamento da fatura do período anterior (t-3);

Rio Grande: Clientes localizados na praça da empresa Rio Grande;

Grãos: Clientes que trabalham com cargas a granel;

Vlr Recebido: Valor da fatura paga pelo cliente;

Garantia Exigida: Clientes de quem a empresa exigiu algum tipo de garantia para efetuar a venda a prazo.

Alguns coeficientes das variáveis apresentam uma relação diferente da esperada. As variáveis “Pag1”, “Pag2” e “Pag3” possuem uma relação positiva com a ocorrência de inadimplência. Isso aponta que clientes que apresentam algum atraso nos pagamentos das 3 últimas faturas provavelmente atrasarão pagamentos futuros. Além disso, há um incremento no valor do coeficiente (beta) na medida em que se comparam pagamentos de faturas mais antigas com pagamentos de faturas mais recentes (Pag1 = 2,173 / Pag3 = 1,351). Isso pode ser indício de uma interação entre as variáveis de comportamento de pagamento.

A variável “Rio Grande”, da mesma forma, apresenta uma relação positiva com a inadimplência. A conclusão é a de que clientes que pertencem à cidade de área de atuação da empresa são mais inadimplentes do que os que pertencem a outras cidades.

A variável “Grãos” indica que os clientes que trabalham com carga do tipo a granel são mais inadimplentes que clientes que trabalham com outros tipos de cargas, uma

explicação possível é que este tipo carga está ligado a questões de safra, sofrendo com períodos de recessão e grande demanda de serviços.

O coeficiente da variável “Vlr Recebido” indica que faturas com valores menores possuem mais chances de se tornarem inadimplentes. Isso pode estar relacionado com o tamanho (em faturamento, número de funcionários, entre outros) do cliente. Outro ponto é que a variável apresenta um coeficiente pequeno (-0,00001146080510504), mas estatisticamente significativo. A critério do autor da pesquisa, manteve-se a variável no modelo, porque ela incrementa sensivelmente a taxa de acerto geral do modelo de risco de inadimplência. Esses resultados serão mais bem descritos mais a frente.

A variável “Garantia Exigida” indica que a empresa consegue perceber empiricamente situações de risco de inadimplência e assim exigir algum tipo de garantia do cliente.

Foi realizada uma série de procedimentos para testar a robustez do modelo de previsão de inadimplência. Inicialmente, foram testadas diferentes configurações de divisão na base de dados entre calibração e projeção, as quais estão apresentadas na Quadro 2.

Divisão %	Calibração	Projeção	Total
80/20	10.893	2.723	13.616
70/30	9.531	4.085	13.616
60/40	8.170	5.446	13.616
50/50	6.808	6.808	13.616

Quadro 2. Configurações testadas na calibração e projeção.

Para verificar a robustez das variáveis, os coeficientes foram estimados utilizando 4 configurações diferentes de quantidade de dados da base. Eles foram estimados no SPSS, utilizando 50% da base, dados de 6.808 faturas. Foram utilizados também 60% com 8.170, 70% com 9.531 e 80% com 10.893. As variáveis em todas as estimações se mantiveram estatisticamente significantes, o que variaram foram os valores dos coeficientes (betas). Esse procedimento originou 4 modelos, cuja única diferença eram os valores dos coeficientes.

Os 4 modelos foram testados utilizando a parte da base reservada para projeção (ver Quadro 2). Os principais critérios adotados para comparação dos modelos foram os testes de *Cox & Snell R Square* e *Nagelkerke R Square*. Para avaliação dos resultados dos modelos, foi utilizada a técnica de *Hit Rate*, que consiste na contagem dos acertos nas classificações corretas das faturas com o auxílio do programa Microsoft Office Excel 2007. A contagem dos acertos no *Hit Rate* utilizou como ponto de corte o “0,5”, segundo Araújo e Carmona (2008). Isso é necessário porque o modelo *logit* tem como saída na classificação da fatura de um

cliente um índice entre zero e um, no qual zero é adimplência, e um é inadimplência. Porém, na realidade, a inadimplência é uma condição binária, ou ela existe ou não (zero ou um). Para as faturas adimplentes (zero) eram considerados acertos se o resultado do modelo fosse menor que 0,5. Para as faturas inadimplentes (um) eram considerados como acertos se o resultado do modelo fosse igual ou maior que 0,5. Os resultados na classificação dos modelos estão no Quadro 3.

Divisão %	Cox & Snell R Square	Nagelkerke R Square	Hit Rate % (0,5)		
			Adimpl.	Inad.	Média
80/20	0,328	0,472	94,79	50,33	72,56
70/30	0,337	0,488	92,22	54,66	73,44
60/40	0,340	0,499	93,51	53,81	73,66
50/50	0,345	0,508	93,69	54,46	74,07

Quadro 3. Resultados dos modelos.

Pode-se observar que os 4 modelos obtiveram resultados semelhantes de classificação das faturas. O melhor modelo foi o que utilizou apenas 50% da base na estimação; isso pode ser observado nos testes de *Cox & Snell R Square* e *Nagelkerke R Square*. Além disso, o modelo 50%/50% também obteve o melhor percentual médio de classificação (74,07%). Outra questão importante é que todos os 4 modelos se revelaram melhores para classificar as faturas adimplentes do que para classificar as faturas inadimplentes. Isso ocorre devido ao desbalanceamento da base de dados, na qual existe maior ocorrência de faturas adimplentes que inadimplentes (Tabela 14).

	Quant.	%	R\$	%
Adimplentes	9.951	73,1	99.748.808	70,6
Inadimplentes	3.665	26,9	41.563.486	29,4
Total	13.616	100	141.312.294	100

Tabela 14. Número de faturas adimplentes e inadimplentes.

É importante ressaltar que em amostras nas quais a proporção de observações de inadimplência e adimplência é igual, ou seja, 50%, basta o modelo obter um percentual de *hit rate* maior que 50% para que este seja considerado útil, dado que mediante classificação aleatória se tem 50% de chance de acerto. Porém para amostras em que existem diferenças estatisticamente significativas na proporção de adimplentes/inadimplentes, isso se torna um problema. Com uma proporção de 73,1% de adimplentes, basta que um modelo classifique

todos os clientes como adimplentes que ele obterá boas taxas de *hit rate* (BARTH, 2004). Desta forma, devem-se ponderar esta medida de utilidade com a fórmula:

$$C = p^2 + (1 - p)^2 \quad (26.1)$$

Onde C é a medida de utilidade do modelo ponderada e p é a proporção de adimplentes na amostra conhecida. Para que o modelo seja considerado útil, ele deve ter uma taxa de acerto (*hit rate*) melhor que 60,6%.

Uma segunda bateria de testes utiliza a técnica de *Jack Knife* para divisão da base de dados em calibração e projeção. Utilizando o Microsoft Office Excel 2007, a base foi organizada em ordem de data de emissão das faturas e, logo após, foi dividida em 5 blocos (A, B, C, D e E). Cada bloco contém 20% do total da base de dados (Tabela 15). A organização da base por data de emissão foi adotada de forma a não haver aglutinação de faturas de um cliente em uma das 5 divisões. As quantidades de ocorrências em cada um dos blocos podem ser demonstradas na Tabela 15.

Bloco (20%)	Quant.	Acumulado
A	2.723	2.723
B	2.723	5.446
C	2.723	8.169
D	2.723	10.892
E	2.724	13.616
Total	13.616	-

Tabela 15. Divisão da base.

Logo após a delimitação dos blocos (utilizando as mesmas 7 variáveis descritas na Tabela 13), estimou-se um modelo para o bloco A e utilizaram-se os blocos B, C, D e E para projeção. Esse procedimento foi aplicado para todos os blocos. Os resultados são apresentados no Quadro 4.

Blocos		Hit Rate % (0,5)			Geral	Cox e Snell R Square	Nagelkerke R Square
Calibração	Projeção	Adimpl.	Inadimpl.	Média			
A (20%)	B	94,91	61,03	77,97	77,65	0,356	0,514
	C	96,08	58,52	77,30			
	D	95,32	61,56	78,44			
	E	96,03	57,73	76,88			
B (20%)	A	95,68	58,52	77,10	76,39	0,345	0,51
	C	96,37	55,17	75,77			
	D	95,48	58,52	77,00			
	E	96,37	55,02	75,69			
C (20%)	B	95,61	60,28	77,95	77,62	0,326	0,485
	C	95,71	59,97	77,84			
	D	95,46	60,13	77,79			
	E	96,78	56,99	76,88			
D (20%)	A	92,50	54,70	73,60	73,37	0,307	0,423
	C	92,05	55,45	73,75			
	D	93,17	53,85	73,51			
	E	93,73	51,50	72,61			
E (20%)	A	94,60	51,47	73,04	72,57	0,259	0,395
	C	94,65	51,96	73,31			
	D	95,60	48,53	72,06			
	E	94,89	48,86	71,87			
Média	-	95,05	55,99		75,52	-	-

Quadro 4. Resultados dos modelos.

Da mesma forma, o ponto de corte foi o 0,5, e o cálculo do *Hit Rate* foi semelhante ao descrito anteriormente. O Quadro 4 mostra que, em todas as situações, os resultados se comportaram de maneira semelhante. O resultado final do método *Jack Knife* é 95,05% de assertividade de classificação nas futuras adimplentes, 55,99% nas futuras inadimplentes e 75,52% na média.

Comparando os resultados deste trabalho com os resultados do estudo de Wilson, Summers e Hope (2000), que também utilizou dados de comportamento de pagamento, na média, o modelo proposto (75,52%) obteve um índice de assertividade superior ao trabalho de Wilson, Summers e Hope (2000) (70,83%). Por outro lado, o modelo de Wilson Summers e Hope (2000) obteve um índice de assertividade melhor para as ocorrências de inadimplência (ver Tabela 3).

Outra forma eficiente de verificar o desempenho do modelo proposto é utilizando as Curvas de Características de Operação do Receptor (*Receiver Operating Characteristic*), mais conhecido como ROC. Além de verificar o desempenho dos modelos, as Curvas ROC são

úteis para seleccionar algoritmos (SPACKMAN, 1989), comparar diferentes modelos concorrentes e demonstrar a relação entre sensibilidade (verdadeiro positivo ou sinal) e especificidade (falso positivo ou ruído) para diferentes pontos de corte (SWETS, 1988; BRAGA, 2000).

No problema de detecção de dois tipos diferentes de situações (adimplência/inadimplência) e duas respostas distintas, existem quatro resultados possíveis. Essa situação é descrita no Quadro 4.1.

		Realidade	
		p	n
Classificação	p	Verdadeiro Positivo Sensibilidade	Falso Positivo Especificidade
	n	Falso Negativo	Verdadeiro Negativo
Total		P	N

Quadro 4.1. Conjunto de respostas possíveis.

A partir do Quadro 4.1, pode ser calculada uma série de métricas, dentre elas, destacam-se a *tp rate* (taxa de acerto do produtor) e *fp rate* (falso alarme do produtor) (FAWCETT, 2005; BRAGA, 2000).

$$tp\ rate = \frac{VerdadeiroPositivo}{P} \quad (27)$$

$$fp\ rate = \frac{FalsoPositivo}{N} \quad (28)$$

Estas duas métricas são utilizadas para a construção do ROC. Os gráficos ROC são bidimensionais. O eixo Y representa o *tp rate*; o eixo X, o *fp rate*. A Figura 9 é uma demonstração de uma Curva ROC.

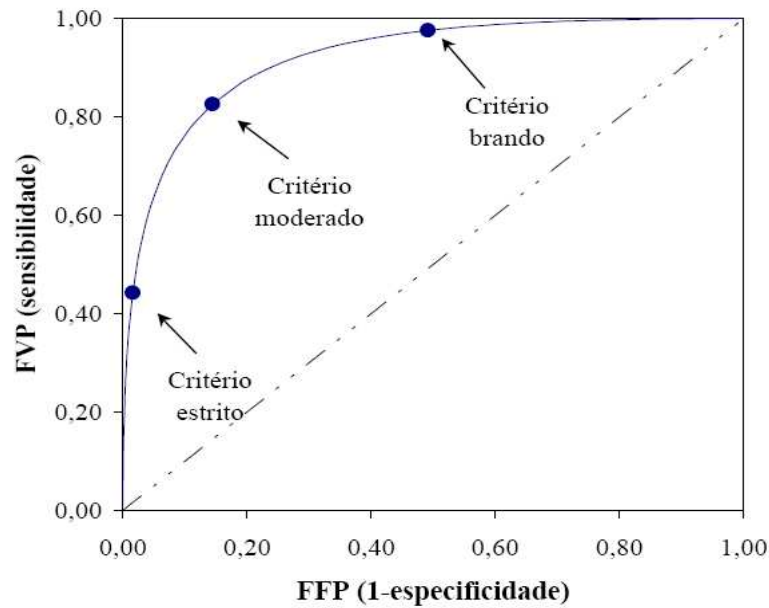


Figura 9. Curva ROC.

Fonte: Braga (2000).

Na Figura 9, a linha em azul representa um modelo de classificação binária. A diagonal pontilhada é uma linha de referência, que representa uma classificação aleatória ou modelo nulo. Curvas localizadas próximas ao canto superior esquerdo do gráfico possuem poder discriminatório melhor que uma classificação aleatória. Curvas localizadas abaixo e à direita possuem um poder discriminatório ruim. Ao longo da curva, podem ser observadas as diferentes configurações entre sensibilidade (verdadeiro positivo) e especificidade (falso positivo), que dependem do critério utilizado. O ponto de corte é um valor definido arbitrariamente pelo pesquisador e define a classificação de uma amostra. Um exemplo é a definição de um ponto de corte de 0,5. Amostras com valor maior ou igual a 0,5 são consideradas como classe um, e amostras com ponto de corte menor que 0,5 são consideradas como classe zero (FAWCETT, 2005). Por outro lado, o ponto de corte adotado depende dos objetivos de cada pesquisador. Na prática, a decisão da adoção de um ponto de corte deve ser pautada nas consequências de se cometer um determinado tipo de erro. Uma empresa com restrições financeiras deve ser mais rigorosa na concessão de crédito, devendo, portanto, adotar um ponto de corte mais estrito (ver Figura 9). Isso diminuiria as chances de se classificar como adimplente um cliente que, na realidade, é/ou será inadimplente.

A Curva ROC permite identificar os valores para otimizar a sensibilidade em função da especificidade. A otimização corresponde ao ponto na curva mais próximo ao canto superior esquerdo, onde o índice de verdadeiros positivos é um (alta sensibilidade) e o índice

de falsos positivos é zero (baixa especificidade). Esse ponto é representado pelo critério moderado na Figura 9 (SWETS, 1988; FAWCETT, 2005).

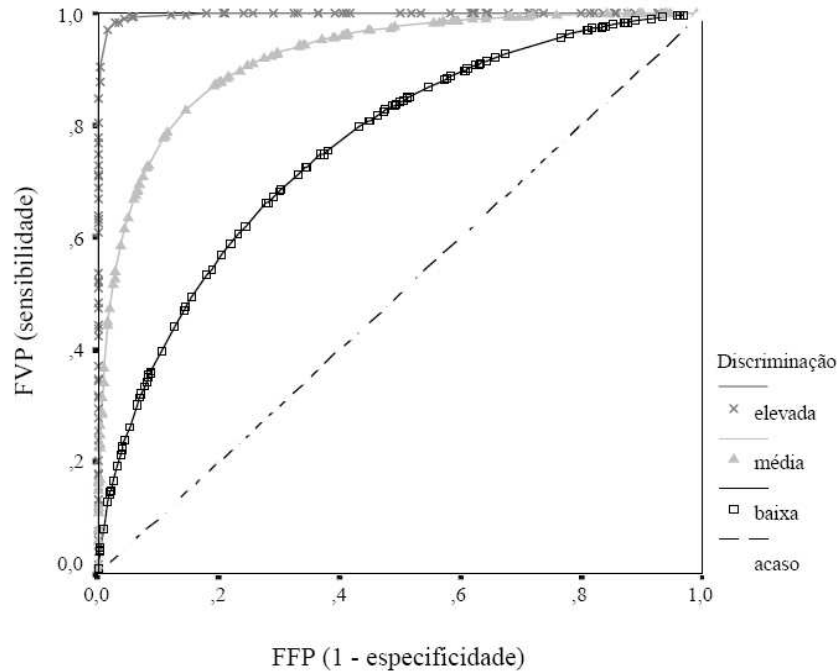
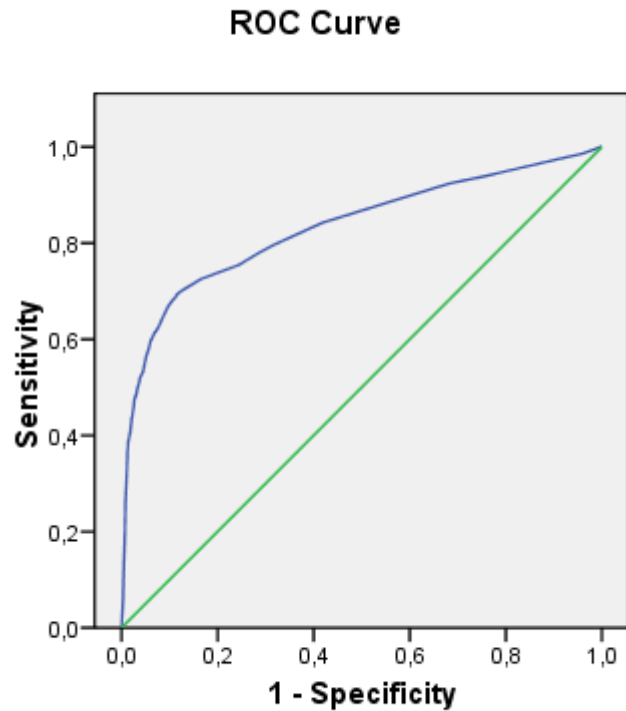


Figura 10. Comparação de modelos com a Curva ROC.

Fonte: Braga (2000).

As Curvas ROC são amplamente utilizadas para comparar o desempenho de modelos de diagnóstico. Na figura 10, podem ser observadas as curvas de três diferentes modelos discriminatórios. Para esta pesquisa, a curva ROC foi elaborada utilizando o SPSS 15. Os resultados da aplicação do teste podem ser observados na Figura 11.



Diagonal segments are produced by ties.

Figura 11. Curva ROC para o Modelo de Risco de Inadimplência.

A área sob a Curva ROC é uma das medidas de desempenho do teste. Um teste sem poder discriminatório tem uma área sob a curva de 0,5 (linha verde). Na medicina, modelos acima de 0,7 são considerados satisfatórios (MAULIK *et al.* 1991). Os resultados da área sob a curva para o modelo proposto estão descritos no Quadro 4.2.

<i>Test Result Variable(s): Completo</i>				
<i>Area</i>	<i>Std. Error(a)</i>	<i>Asymptotic Sig.(b)</i>	<i>Asymptotic 95% Confidence Interval</i>	
<i>Lower Bound</i>	<i>Upper Bound</i>	<i>Lower Bound</i>	<i>Upper Bound</i>	<i>Lower Bound</i>
0,833	0,004	0,000	0,824	0,842

Quadro 4.2. Resultados Curva ROC.

A significância estatística é inferior a 0,05 (0,00), o que aponta a validade do modelo. A área sob a curva 0,833 é um resultado estatístico útil para determinar a exatidão do modelo. Na Tabela 15.1 são observadas as sensibilidades e as especificidades para diferentes pontos de corte.

Ponto de corte	Sensibilidade	1 - Especificidade	Ponto de corte	Sensibilidade	1 - Especificidade
0,06	1,00	1,00	0,50	0,57	0,05
0,07	1,00	1,00	0,51	0,57	0,05
0,08	1,00	0,99	0,52	0,56	0,05
0,09	1,00	0,99	0,53	0,55	0,05
0,10	0,99	0,96	0,54	0,54	0,05
0,11	0,94	0,76	0,55	0,54	0,05
0,12	0,92	0,68	0,56	0,54	0,05
0,13	0,84	0,42	0,57	0,53	0,04
0,14	0,80	0,32	0,58	0,53	0,04
0,15	0,78	0,29	0,59	0,52	0,04
0,16	0,75	0,24	0,60	0,52	0,04
0,17	0,73	0,17	0,61	0,52	0,04
0,18	0,70	0,12	0,62	0,51	0,04
0,19	0,69	0,12	0,63	0,50	0,03
0,20	0,69	0,11	0,64	0,49	0,03
0,21	0,69	0,11	0,65	0,49	0,03
0,22	0,68	0,11	0,66	0,49	0,03
0,23	0,68	0,11	0,67	0,49	0,03
0,24	0,68	0,11	0,68	0,48	0,03
0,25	0,68	0,10	0,69	0,48	0,03
0,26	0,68	0,10	0,70	0,48	0,03
0,27	0,67	0,10	0,71	0,48	0,03
0,28	0,67	0,10	0,72	0,48	0,03
0,29	0,66	0,09	0,73	0,47	0,03
0,30	0,66	0,09	0,74	0,47	0,03
0,31	0,65	0,09	0,75	0,46	0,03
0,32	0,65	0,09	0,76	0,45	0,02
0,33	0,64	0,08	0,77	0,44	0,02
0,34	0,63	0,08	0,78	0,44	0,02
0,35	0,62	0,07	0,79	0,42	0,02
0,36	0,62	0,07	0,80	0,41	0,02
0,37	0,62	0,07	0,81	0,40	0,02
0,38	0,62	0,07	0,82	0,40	0,02
0,39	0,62	0,07	0,83	0,40	0,02
0,40	0,61	0,07	0,84	0,40	0,02
0,41	0,61	0,07	0,85	0,39	0,01
0,42	0,60	0,06	0,86	0,39	0,01
0,43	0,60	0,06	0,87	0,38	0,01
0,44	0,60	0,06	0,88	0,37	0,01
0,45	0,60	0,06	0,89	0,37	0,01
0,46	0,59	0,06	0,90	0,36	0,01
0,47	0,59	0,06	0,91	0,33	0,01
0,48	0,57	0,05	0,92	0,26	0,01
0,49	0,57	0,05	0,93	0,20	0,01

Tabela 15.1. Pontos de corte Curva ROC.

A Tabela 15.1 demonstra o comportamento da sensibilidade e especificidade para os pontos de corte de 0,06 a 0,93. A sensibilidade representa a exatidão do modelo de risco de inadimplência em detectar clientes inadimplentes quando de fato o são. A especificidade é a proporção de clientes adimplentes classificados como inadimplentes (erro tipo I). No ponto de corte 0,49, a sensibilidade é 0,57 e a especificidade é 0,05. Isso significa que, aproximadamente, 57% de todos os clientes inadimplentes seriam corretamente classificados, enquanto que 5% dos clientes adimplentes poderiam ser incorretamente classificados como inadimplentes.

Foram testadas outras variações do modelo de risco de inadimplência com o objetivo de verificar a capacidade discriminante das diferentes classes de variáveis utilizadas. Os resultados podem ser observados na Figura 11.1 e Tabela 15.2.

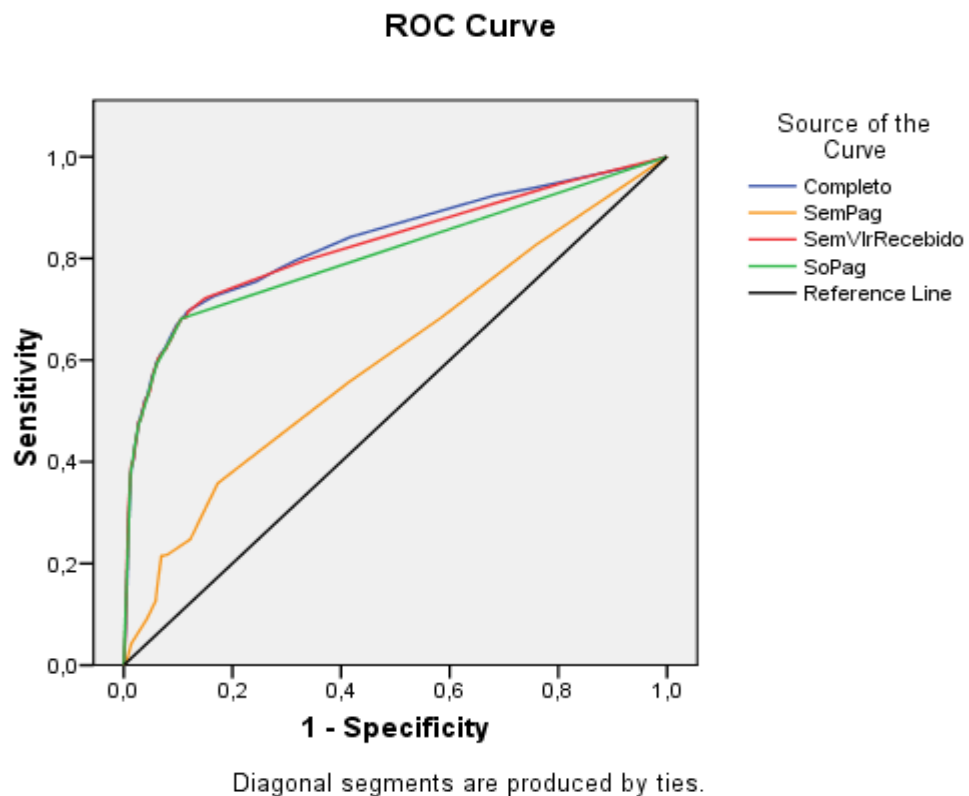


Figura 11.1. Comparação de modelos utilizando a curva ROC.

Os modelos comparados foram:

- Completo: é o modelo utilizado na pesquisa e descrito na Tabela 11. Ele é composto por todas as variáveis, sendo: “Pag1”, “Pag2”, “Pag3”, “Rio Grande”, “Vlr recebido”, “Garantia exigida” e “Grãos”;
- Sem Pag: Modelo sem as variáveis de comportamento de pagamento. Ele é composto pelas variáveis: “Rio Grande”, “Vlr recebido”, “Garantia exigida” e “Grãos”;
- Sem Vlr Recebido: Modelo sem a variável valor da fatura (“Vlr recebido”). Ele é composto pelas variáveis: “Pag1”, “Pag2”, “Pag3”, “Rio Grande”, “Garantia exigida” e “Grãos”;
- Só Pag: Modelo que utiliza apenas as variáveis de comportamento de pagamento, sendo: “Pag1”, “Pag2” e “Pag3”.

Modelos	Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound
Completo	0,833	0,004	0,000	0,824	0,842
SemPag	0,601	0,006	0,000	0,590	0,612
SemVlrRecebido	0,827	0,005	0,000	0,817	0,836
SoPag	0,807	0,005	0,000	0,797	0,817

Quadro 4.3. Comparação de modelos utilizando a curva ROC.

Como observado na Tabela 4.2 e 4.3, o modelo de risco de inadimplência, identificado aqui como “Completo”, possui uma área sob a curva de 0,833. O modelo “Só Pag” composto apenas por variáveis de comportamento de pagamento mostra que essas variáveis possuem grande poder discriminatório e detêm 0,807 da área sob a curva. As demais variáveis representadas pelo modelo “Sem Pag” possuem área sob a curva de 0,601. Esse modelo não apresenta resultados satisfatórios segundo critério de MAULIK *et al.* (1991), que afirmam serem os modelos com área a partir de 0,7 considerados bons.

O modelo “Sem Vlr recebido” foi testado porque o coeficiente da variável “Vlr Recebido” se mostrou baixo (-0,00001146080510504), porém significativo estatisticamente. Por esse motivo, foi comparado o desempenho do modelo sem a adição desta variável. Na Figura 11.1 e no Quadro 4.3, observa-se que a adição desta variável melhora sensivelmente o modelo, passando de uma área sob a curva de 0,827 para 0,833. No Apêndice D estão descritos os pontos de corte para cada modelo testado.

Depois de concluídos os testes da capacidade preditiva do modelo de risco de inadimplência (*Hit rate*, *Hit rate Jackknif* e ROC), os resultados do modelo de risco de

inadimplência para os principais clientes em termos de valor de faturamento da empresa são descritos na Tabela 16.

Cliente	Faturamento total (2008/2009)	Inadimplência média
49 e 1568	18.669.158,60	0,03
605 e 844	12.084.928,85	0,30
144	10.508.034,23	0,25
245	8.483.218,86	0,04
594	7.737.795,77	0,05
145	5.295.739,82	0,89
589	5.214.340,43	0,61
Outros	73.319.077,75	0,12
Média Geral	141.312.294,31	0,29

Tabela 16. Média do risco de inadimplência para os clientes.

Os índices de inadimplência médio de cada cliente são calculados, fazendo uma média simples dos índices de inadimplência de todas as faturas pagas por aquele mesmo cliente. A interpretação dos resultados do modelo de inadimplência é: quanto mais perto de 1, maior o risco de inadimplência o cliente possui, e quanto mais próximo de zero, menor o risco de inadimplência. Na Tabela 12, estão numa situação de pouco risco de inadimplência os clientes “49e1568”, “245”, “594” e “Outros”. Estão numa situação intermediária de risco de inadimplência os clientes “605e844” e “144”. Estão numa situação de alto risco os clientes “589” e “145”. A média geral de todos os clientes da empresa é 0,29. É importante ressaltar que estes sete clientes são os que possuem o maior faturamento.

Cliente	Número de faturas	Faturas inadimplentes	Média de Risco
49e1568	749	94	0,03
605e844	555	235	0,30
144	669	245	0,25
245	639	97	0,04
594	331	26	0,05
145	198	198	0,89
589	173	150	0,61
Outros	10.302	2.620	0,12
Geral	13.616	3.665	0,29

Tabela 17. Média do Risco de inadimplência e número de faturas emitidas por cliente.

Na Tabela 17, estão os sete principais clientes da empresa estudada e seus respectivos riscos médios. Os riscos apresentados na Tabela 13 são uma média dos riscos de todas as

faturas pagas de cada cliente. Na linha “Outros”, está acumulado o restante dos clientes da empresa estudada. A coluna “Número de faturas” é uma contagem de todas as faturas emitidas. A coluna “Faturas inadimplentes” é o número de faturas pagas com atraso. O cliente “605e844” teve um total de 555 faturas. Deste total, 235 faturas foram pagas em atraso. A média do índice de risco de inadimplência das 555 faturas é 0,30. Na Figura 12, estão dispostos, em sequência cronológica, os resultados do modelo de inadimplência para as faturas pagas pelo cliente “605e844”.

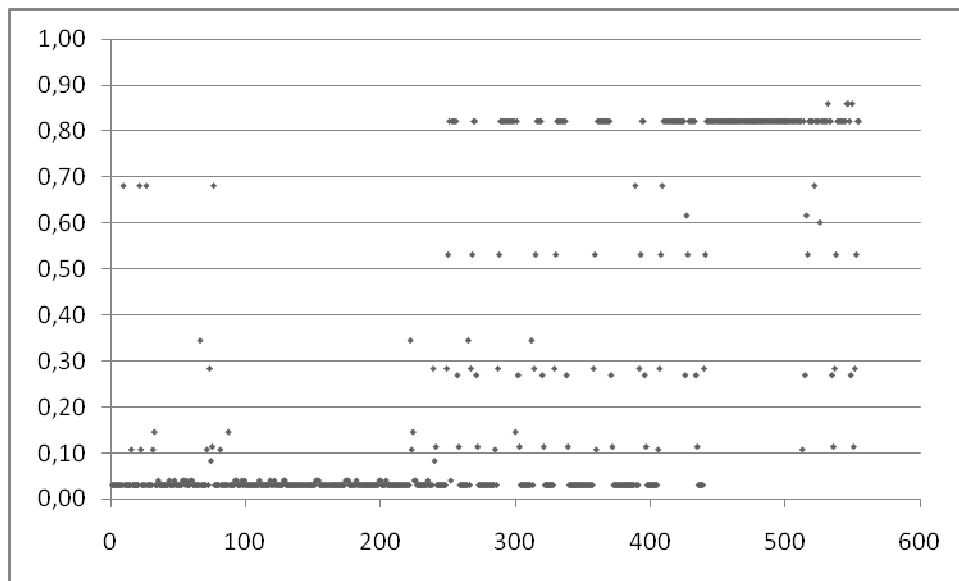


Figura 12. Evolução do risco de inadimplência estimado para o cliente “605e844”.

No Apêndice C, estão os gráficos dos riscos de inadimplência dos demais clientes. Na próxima seção será abordada a forma como se deu a inclusão destes resultados no modelo de CLV.

4.2. MODELO PROPOSTO

Após estruturar o modelo de inadimplência, é necessário integrar seus resultados ao modelo de CLV. A ideia do modelo de inadimplência é a de “punir” as receitas de um cliente que possui maior risco. Esse mecanismo é descrito de maneira simplificada no Quadro 5.

Cliente	Receita	Risco de inadimplência	Adequação do risco de inadimplência	Receita com risco de inadimplência
A	200	0,60	$- 0,60 + 1 = 0,40$	$200 * 0,40 = 80$
B	100	0,15	$- 0,15 + 1 = 0,85$	$100 * 0,85 = 85$
C	90	0,01	$- 0,01 + 1 = 0,99$	$90 * 0,99 = 89,1$

Quadro 5. Procedimento para integração do modelo de risco de inadimplência ao modelo de CLV.

O quadro 5 exemplifica um passo a passo de como os resultados do modelo de inadimplência foram integrados ao modelo de CLV. Na primeira coluna, temos 3 clientes (A, B e C) em ordem crescente de receita (200, 100 e 90, respectivamente). Ao lado da coluna da Receita está o risco de inadimplência de cada cliente. O índice de risco de inadimplência é interpretado da seguinte forma: quanto mais próximo de 1, mais risco de inadimplência possui o cliente, e quanto mais perto de zero, menos risco de inadimplência. Não seria indicado simplesmente multiplicar o índice do risco de inadimplência com a receita, pois os clientes com menor risco de inadimplência teriam o seu valor de receita mais punido comparado aos clientes que possuem maior risco de inadimplência. Em um caso extremo, um cliente com receita de 100 e índice de risco de inadimplência de 0,30 (situação de baixo risco) teria sua receita reduzida para 30. Para corrigir isso, foi necessário calcular o seu oposto, diminuindo o índice de risco de inadimplência de 1. Detalhes deste cálculo são demonstrados na coluna “Adequação do risco de inadimplência”, no Quadro 5. Este procedimento inverte a interpretação do índice para: quanto mais perto de zero, mais risco de inadimplência, e quanto mais perto de um, menos risco de inadimplência. Logo após a adequação do índice de inadimplência, deve-se multiplicar a receita e o índice de risco. Os detalhes deste cálculo são demonstrados na última coluna do Quadro 5. Após a multiplicação, é obtida a nova receita de cada cliente, a qual é introduzida ao modelo de CLV.

O horizonte de cálculo do CLV é o de 12 meses (ano de 2010). Esse horizonte é justificado por dois motivos: a empresa estudada considera este prazo suficiente devido ao mercado de atuação, semelhante ao procedimento utilizado por Kumar *et al.* (2008). O segundo motivo é porque os termos acumulados do CLV tendem a estabilizar no longo prazo devido à incidência de uma alta taxa de desconto. Quanto mais elevadas forem as taxas de desconto, menos tempo o CLV acumulado leva para estabilizar (Figura 13).

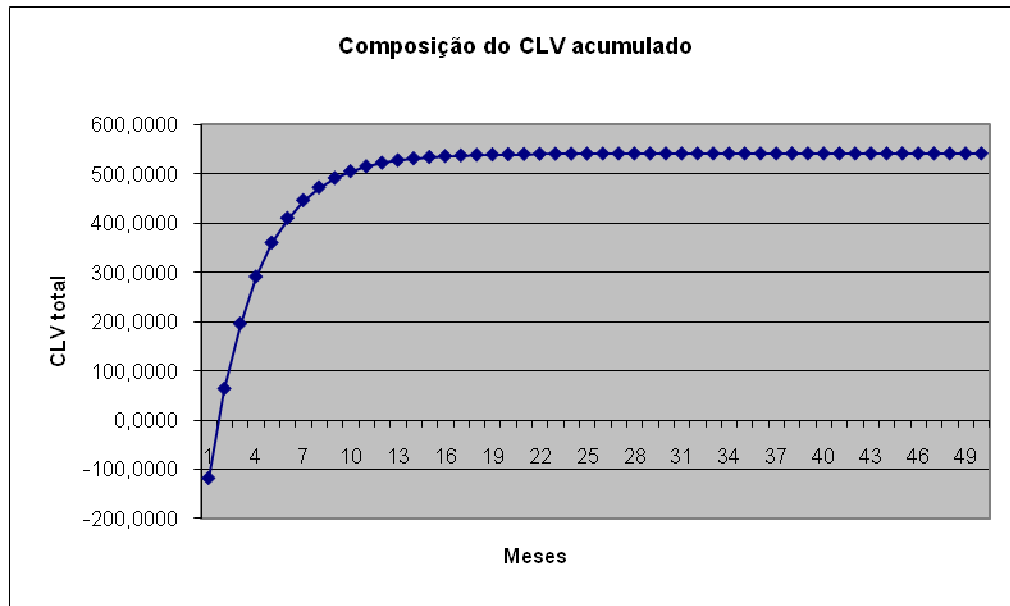


Figura 13. Curva dos valores dos termos do CLV.

Se for adotada como taxa de desconto no modelo (i) a taxa de juros brasileira, e for comparado o contexto brasileiro (taxa 10% a.a.) e o americano (0% a 0, 25% a.a.), espera-se que os termos acumulados do CLV no contexto brasileiro tendam a estabilizar-se mais rápido, se comparados ao contexto americano. Neste estudo, a taxa de desconto adotada é o custo de capital para a empresa, o qual foi obtido com leitura de relatórios financeiros e bancários da empresa estudada.

Os preços de venda e de compra do diesel foram coletados dos relatórios de vendas e notas fiscais de compras das filiais da empresa estudada. O preço de venda é o mesmo para todos os postos da rede (R\$ 2,2825). O preço de compra foi obtido utilizando uma média simples dos preços de compra dos postos de rodovia (R\$ 1,9275). Os preços de compra foram coletados em janeiro de 2009, a partir das notas fiscais de compra das distribuidoras (Tabela 18).

Distribuidor	Preço de compra	Prazos de compra (dias)
BR	1,9012	28
Shell	1,9172	15
Esso	1,9403	7
Ipiranga	1,9511	10
Média	1,9275	15

Tabela 18. Preços de compra do diesel.

O prazo de compra do diesel (15 dias) também é uma média simples dos prazos de todos os fornecedores (Petrobras, Shell, Ipiranga e Esso). O prazo de recebimento do cliente é descrito no seu cadastro individual, obtido no setor de faturamento da empresa estudada. A quantidade de diesel futura demandada por cada cliente foi obtida com uma média simples das compras de 2008 e 2009, procedimento que foi adotado por CAMARGO (2009). No lugar da média, uma alternativa seria adotar a técnica de previsão de séries temporais, porém, mais que aperfeiçoar o método e os resultados finais, o principal objetivo do trabalho de campo desta pesquisa é verificar a aplicabilidade do modelo proposto.

O cálculo da quantidade futura demandada é a média dos litros comprados pelos clientes (ver Apêndice A). As receitas foram obtidas multiplicando-se os litros com o preço de venda (R\$ 2,2825). Os custos são obtidos multiplicando-se os litros com o preço de compra (R\$ 1,9275). O risco de inadimplência futuro de cada cliente (μ_j) será aqui considerado constante e é uma média do risco de inadimplência de todas as faturas na base e pagas pelo cliente. Logo após, os dados foram inseridos no modelo (29) e o CLV mensurado para os 7 principais clientes em valores de faturamento da empresa estudada.

$$CLV_j = \sum_{t=1}^T q_j \left[\mu_j P \frac{1-d_j}{(1+i)^{s_j}} - \frac{C}{(1+i)^z} \right] \quad (29)$$

Assim:

j = cliente;

t = tempo;

CLV_j = valor do cliente j ;

q_j = litros de combustíveis adquiridos pelo cliente j ;

P = preço do litro de combustível pago pelo cliente;

C = preço do litro de combustível pago ao fornecedor;

s_j = prazo concedido ao cliente j ;

z = prazo cedido pelo fornecedor;

μ_j = risco de inadimplência do cliente j ;

d_j = desconto concedido ao cliente j ;

Os resultados do modelo são mostrados no Quadro 6.

Cliente j	Inadimplência η_j	Desconto d	Prazo S	Receita $q_j P$	Custo $q_j C$	CLV
49e1568	0,97	9%	7	358.618,26	302.841,93	150.290,99
605e844	0,70	5%	15	316.097,57	266.934,53	-603.061,04
144	0,75	5%	45	205.385,80	173.441,89	-327.945,85
245	0,96	7%	15	167.276,15	141.259,48	85.338,57
594	0,95	5%	32	153.024,72	129.224,60	68.894,84
145	0,11	3%	20	107.640,66	90.899,18	-828.315,68
589	0,39	5%	20	105.247,56	88.878,28	-518.293,95
Outros	0,88	3%	15	1.483.188,67	1.252.506,53	184.957,92
CE						-1.788.134,20

Quadro 6. Resultados do modelo proposto.

Os clientes “605e844”, “144”, “145” e “589” possuem um CLV negativo. Os clientes “49e1568”, “245”, “594” e “Outros” apresentaram um CLV positivo. No geral, a carteira de clientes que trabalha com ordem de frete apresenta um *Customer Equity* (CE, somatório do CLV dos clientes) negativo. Na Tabela 19, pode ser observada a relação do CLV com o risco de inadimplência para o cliente “49e1568”.

Risco de inadimplência	CLV
0	-3.136.333,25
0,1	-2.796.103,00
0,2	-2.455.872,75
0,3	-2.115.642,49
0,4	-1.775.412,24
0,5	-1.435.181,99
0,6	-1.094.951,74
0,7	-754.721,48
0,8	-414.491,23
0,9	-74.260,98
1	265.969,28

Tabela 19. Relação do valor do cliente com o risco de inadimplência.

O risco do cliente “49e1568” é 0,97. Na Tabela 19, foram simulados outros níveis de risco de inadimplência. A Tabela mostra que, para este cliente, um risco de inadimplência próximo de 0,9 é o bastante para tornar o CLV do cliente negativo. Essa relação descrita na Tabela 19 é representada graficamente na Figura 14.

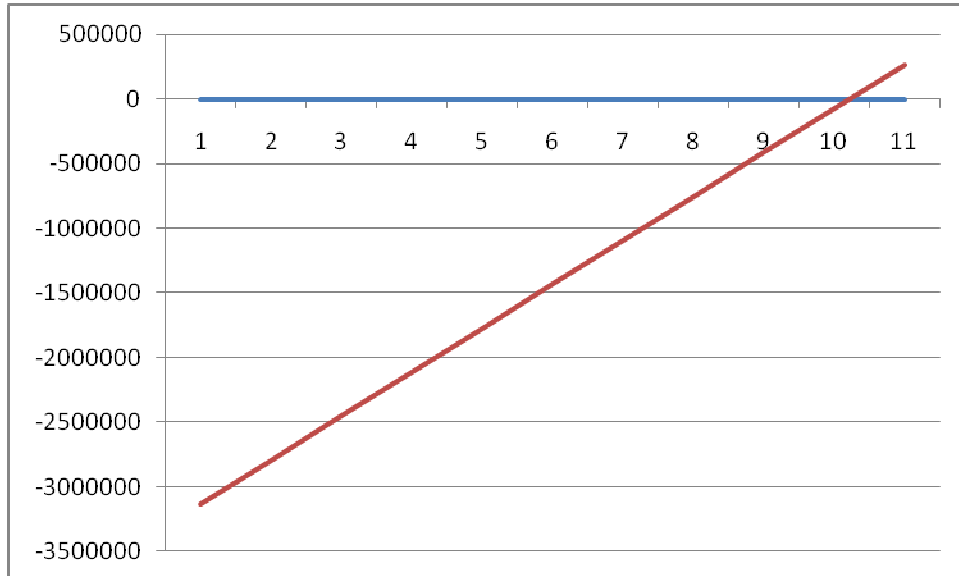


Figura 14. Relação do valor do cliente com o risco de inadimplência.

No caso de risco de inadimplência ser zero, a empresa poderá arcar com prejuízos altos, tendo em vista que terá de pagar os custos, mas não receberá as receitas. A partir da Figura 14, verifica-se que existe uma relação inversamente proporcional entre o risco de inadimplência e o CLV. A seguir serão apresentadas as considerações finais do trabalho.

5. CONSIDERAÇÕES FINAIS

Neste capítulo serão discutidas as contribuições deste estudo, as implicações gerenciais dos resultados, as limitações do trabalho e também sugestões para estudos futuros. A realização deste trabalho contribuiu com uma série de ideias, algumas inéditas, tanto para a discussão sobre mensurar o valor do cliente, quanto na discussão em relação a modelos de risco de inadimplência.

Com relação à construção de modelos de inadimplência, este estudo elaborou um modelo capaz de estimar o risco de inadimplência para cada pagamento realizado por um cliente. Isso sugere que o risco de inadimplência de um cliente não é fixo, mas sim varia no tempo. Os modelos normalmente encontrados na literatura mensuram apenas um risco para o cliente (BEAVER, 1966; ALTMAN 1968; SANVICENT e MINARDI 1979; RIBEIRO, 2008). Outro ponto é que a definição de inadimplência utilizada nesta pesquisa é mais adequada ao contexto de risco de crédito comercial. Normalmente os modelos de previsão de inadimplência são elaborados para o contexto de crédito bancário. As principais diferenças entre os dois é que no contexto de crédito bancário há a existência de variáveis independentes no modelo, que são originadas de análises de balanço, e a definição de risco de inadimplência é a insolvência do cliente (BEAVER, 1966; ALTMAN 1968; SANVICENT e MINARDI 1979). Já os modelos realizados para o contexto de crédito comercial são escassos na literatura (GITMAN, 2008). Esses modelos se diferem porque utilizam como definição de risco de inadimplência um prazo (normalmente número de dias) de atraso no pagamento. Outra diferença é que normalmente este tipo de modelo utiliza variáveis independentes sobre o perfil socioeconômico do cliente (RIBEIRO 2008).

Uma segunda contribuição teórica desta pesquisa é a adequação do modelo básico de CLV para o contexto de venda a prazo e de assimetria entre os prazos de pagamento dos custos e recebimento das receitas.

A terceira contribuição teórica é a discussão realizada sobre a integração dos conceitos de valor do cliente e de risco de inadimplência. Mas não só a discussão realizada, mas também o método e as técnicas utilizadas para efetivamente integrar os conceitos em um modelo matemático. A consecução do estudo de campo deu mais robustez ao modelo, além de contribuir com a realização deste tipo de estudo, que é escasso na literatura (KUMAR *et al.* 2008).

5.1. IMPLICAÇÕES GERENCIAIS

Na empresa estudada, o ranking dos clientes é realizado utilizando o valor do faturamento mensal e é utilizado principalmente para direcionar em quais clientes a empresa deve focar. Ela costuma conceder melhores níveis de desconto para empresas com maior faturamento.

Ela também realiza uma avaliação empírica do risco dos clientes. Essa avaliação é realizada sem a ajuda de métodos estruturados, e o resultado acaba sendo condicionado à percepção dos administradores, que nem sempre detêm conhecimento total da situação dos clientes. Um ranking realizado em agosto de 2008, a partir do total de faturamento, tinha como primeiros classificados os clientes apresentados na Tabela 20.

Ranking	Cliente	Valor do faturamento
1º	49e1568	18.669.158,60
2º	605e844	12.084.928,85
3º	144	10.508.034,23
4º	245	8.483.218,86
5º	594	7.737.795,77
6º	145	5.295.739,82
7º	589	5.214.340,43
-	Outros	73.319.077,75
-	Total	141.312.294,31

Tabela 20. Ranking realizado pela empresa estudada.

O cliente “49e1568” aparece em primeiro lugar, seguido pelos clientes “605e844” e “144”. O cliente “Outros” é um somatório do restante dos clientes. No entanto, a adoção de métodos estruturados pode contribuir para aperfeiçoar e complementar os métodos existentes de ranqueamento, permitindo à empresa avaliar seus clientes por novos ângulos. Na Tabela 21, está uma sugestão de ranking baseada nos resultados do modelo proposto.

Ranking	Cliente	CLV
1º	49e1568	222.524,49
6º	605e844	-603.061,04
4º	144	-192.993,98
2º	245	129.205,02
3º	594	111.740,36
7º	145	-828.315,68
5º	589	-518.293,95
-	Outros	591.854,98
-	CE	-1.087.339,81

Tabela 21. Ranking sugerido pela pesquisa.

Comparando os dois rankings, pode ser observado que o cliente “49e1568” é primeiro em ambos. Porém a ordem de classificação muda para o restante dos clientes. O cliente “605e844”, que aparece em segundo na classificação realizada pelo faturamento (Tabela 20), aparece em sexto lugar na classificação realizada pelo CLV (Tabela 21). No entanto, resultados tão conflitantes indicam que estas duas classificações são formas diferentes de avaliar os clientes e, desse modo, não podem ser julgadas como melhor ou pior. Para uma empresa com objetivos de longo prazo seria indicado utilizar o ranking pelo CLV (Tabela 21), pois a visão do resultado das operações é também a de longo prazo.

Por outro lado, empresas com objetivos de curto prazo podem optar por um ranking baseado no faturamento (Tabela 20). Não é um cenário incomum o fato de as empresas se utilizarem de financiamentos de curto prazo (desconto de títulos) para efetuar operações financeiras e saldar dívidas. O cliente “605e844” é um bom exemplo, pois ele possui o segundo lugar no ranking em faturamento (curto prazo), porém possui resultados ruins a longo prazo (CLV = R\$ -603.061,04). Outra questão é com relação ao compromisso de venda do posto com as companhias (Petrobras, Shell, Esso e Ipiranga). Se a empresa se focar apenas em clientes rentáveis poderá haver multas pelo não-cumprimento das metas de vendas.

Estes resultados foram analisados pela empresa e, no início de 2010, houve uma série de mudanças. É importante ressaltar que os resultados do modelo não foram decisivos para que ocorressem essas mudanças, mas sim corroboraram para ajudar a empresa a detectar pontos de melhorias na base de clientes e assim ajudar a redefinir algumas questões, tais como: (a) definição de descontos, (b) definição de prazos de pagamento e mudanças nas políticas promocionais para clientes. As principais mudanças ocorridas estão apresentadas no Quadro 7.

Cliente j	Inadimplência η_j	Desconto d	Prazo S	Receita $q_j P$	Custo $q_j C$	CLV
49e1568	0,97	7%	7	358.618,26	302.841,93	222.524,49
605e844	0,70	5%	15	Rejeitado	Rejeitado	Rejeitado
144	0,75	0%	1	205.385,80	173.441,89	-192.993,98
245	0,96	5%	7	167.276,15	141.259,48	129.205,02
594	0,95	3%	21	153.024,72	129.224,60	111.740,36
145	0,11	3%	20	Rejeitado	Rejeitado	Rejeitado
589	0,39	5%	20	Rejeitado	Rejeitado	Rejeitado
Outros	0,88	0%	15	1.483.188,66	1.252.506,53	591.854,98
CE						862.330,86

Quadro 7. Mudanças ocorridas na empresa estudada.

Comparando o Quadro 7 com o Quadro 6 (pág. 77), pode-se ter uma ideia melhor das diversas mudanças ocorridas na empresa estudada, sendo que as principais foram:

- O cliente “49e1568” mudou o desconto concedido de 9% para 7%;
- No início de 2010, o cliente “605e844” foi rejeitado devido a uma série de atrasos nos pagamentos. Até o momento, os valores não tinham sido pagos;
- O cliente “144” teve uma série de atrasos e renegociou sua dívida. Atualmente está realizando compras na empresa estudada, porém só com pagamento à vista. Ele está efetuando o pagamento parcelado de sua antiga dívida. Houve uma redução no seu prazo de pagamento de 45 dias para 1 dia e foi retirado seu desconto;
- Há na empresa estudada uma discussão entre os proprietários sobre rejeitar ou não o cliente “144”. Sua dívida aproximada é R\$ 1.700.00,00. Até maio de 2010, houve um acréscimo em sua dívida de R\$ 147.000,00, sendo este um valor próximo ao CLV previsto do cliente “144” para 2010 (Quadro 7);
- O cliente “245” efetuou uma redução na taxa de desconto de 7% para 5% e uma redução no prazo de 15 dias para 7 dias;
- O cliente “594” teve mudanças no prazo (de 32 dias para 21 dias) e na taxa de desconto (5% para 3%);
- Os clientes “145” e “589” foram demitidos. Ambos estão sendo cobrados judicialmente;
- Se forem comparadas as duas situações descritas no Quadro 6 e no Quadro 7, as mudanças realizadas na empresa estudada podem resultar em um ganho estimado de R\$ 2.650.465,06 para o ano de 2010.

É esperado que uma série de consequências deva surgir devido às mudanças na política de descontos e prazos de pagamento, descritas no Quadro 7. A empresa tem expectativa de que aconteçam efeitos negativos no curto e médio prazo e que bons resultados virão apenas no longo prazo (dois anos ou mais). Apesar de este trabalho ter sido concluído em julho de 2010 e as mudanças terem ocorrido no início do ano (fevereiro e março), foram observadas algumas consequências. Como foram rejeitados clientes que tinham considerável volume de participação nas vendas, o primeiro efeito foi a diminuição da necessidade de capital para efetuar o troco para as ordens de frete.

A diminuição no volume de vendas também causou uma diminuição nos recebíveis da empresa, que antecipa boa parte de seus recebíveis para saldar seus compromissos e financiar o capital de giro. Porém esta operação se mantém inalterada, pois já há algum tempo a empresa não conseguia antecipar duplicatas dos clientes rejeitados (“605e844”, “145” e “589”) devido à existência de restrições cadastrais nos bancos e instituições financeiras. A forte redução nas vendas também forçou a empresa a pensar em demissões, porém a inauguração de duas unidades de postos de cidade fez com que ela redistribuísse os funcionários de forma que não houvesse demissões.

Devido à rejeição de diversos clientes com expressiva venda de diesel, espera-se que os postos de rodovia não atinjam suas metas mensais de venda com as companhias. Isso pode acarretar multas ou aumento no prazo dos contratos com os distribuidores (Petrobras, Shell, Esso e Ipiranga). Por outro lado, ocorreu uma série de fatores que acarretou um aumento nas vendas dos postos de cidade, principalmente na gasolina, contribuindo para diminuir os efeitos negativos das mudanças nos postos de rodovia.

A empresa estudada espera que as reduções nas taxas de desconto e reduções nos prazos de recebimento dos clientes causem diminuição nas vendas com ordem de frete. Porém o modelo não contempla essa situação, sendo esta uma de suas limitações. Uma forma de contornar este problema é incluir um coeficiente de elasticidade. A seguir, serão discutidas as limitações e sugestões para trabalhos futuros no intuito de melhorar o modelo proposto.

5.2. LIMITAÇÕES E TRABALHOS FUTUROS

A fase de aplicação prática do trabalho proporcionou uma série de novas ideias que podem tanto suprir limitações quanto propor melhorias no modelo proposto.

- No trabalho, foi utilizado um ponto de corte para cálculo do *Hit Rate* de 0,5. Uma alternativa seria testar outros pontos de corte através da Curva ROC e escolher uma que maximize a eficiência em produzir resultados corretos nas classificações do modelo, assim como utilizado no trabalho de Ribeiro (2008);
- O desbalanceamento da base de dados com relação ao número de adimplentes e inadimplentes pode influenciar os resultados de classificação do modelo. Uma alternativa seria utilizar bases de dados mais equilibradas para a estimação do modelo de risco de inadimplência. Uma sugestão seria utilizar o método conhecido como *oversampling* que consiste em igualar a proporção de adimplentes e inadimplentes na amostra utilizada para estimar o modelo. Uma segunda sugestão seria utilizar a fórmula de Anderson (1982) para corrigir o b_0 caso se mantiver proporções diferentes entre adimplentes e inadimplentes;
- Foram testados dois tipos de métodos de estimação nesta pesquisa: *forward LR* e *backward LR*. Embora os resultados dos modelos *backward LR* não tenham sido incluídos nesta pesquisa, em geral, os modelos de risco de inadimplência estimados com este método obtiveram taxas de acerto (*hit rate*) sensivelmente melhor quando comparados ao método *forward LR*. Por outro lado, o método *backward LR* mostrou-se mais instável. Dentre os problemas observados destacam-se: (a) a inclusão de variáveis com significância maior que 0,05 no modelo e (b) os modelos incluíam ou excluíam diversas variáveis de acordo com a configuração de dados utilizados na projeção (50%/50%, 60%/40%, 70%/30% e 80%/20%);
- Nesta pesquisa foram testadas outras duas definições de inadimplência, sendo: (a) atrasos iguais ou superiores a 15 dias e (b) atrasos iguais ou superiores a 30 dias. Embora estes resultados não tenham sido incluídos na pesquisa, o modelo que utilizou a definição de 15 dias obteve taxas de *hit rate* piores comparadas ao modelo que utilizou a definição de sete dias. Já o modelo que utilizou a definição de 30 dias obteve taxas de classificação melhores. Aqui se destacou o modelo que utilizou as seguintes

configurações: (a) definição de inadimplência de 30 dias ou mais, (b) configuração de divisão da base de 80% para estimação e 20% para projeção e (c) método de estimação *forward LR*, obtendo assim uma taxa média de *hit rate* de 100%;

- Para o modelo de CLV, os dados futuros utilizados foram as médias dos dados passados. Admitir que o futuro seja uma repetição do passado não é uma ideia robusta, e atualmente existem diversas técnicas de previsão, tais como regressão linear e técnicas de séries temporais que possivelmente podem prever o futuro com um grau mais apurado de certeza. Uma sugestão para os trabalhos futuros seria adotar tais técnicas para prever as quantidades demandadas pelos clientes, o seu risco de inadimplência futura, preços de compra e de venda do diesel, taxas de desconto/juros e até mesmo a quantidade de clientes existente na carteira da empresa estudada;
- No cálculo do CLV, o risco de inadimplência futuro de cada cliente foi considerado constante (média do passado). Seria interessante considerar no modelo o risco de inadimplência variável;
- Este trabalho utilizou dados originados do cadastro dos clientes e seus dados de pagamento passado. É comum em modelos de previsão de inadimplência utilizar dados de análise de balanço (índice de liquidez, entre outros). Neste trabalho não foi possível obter acesso a dados de balanço dos clientes, porém seria interessante testar variáveis de análise de balanço com dados de comportamento de pagamento semelhante ao realizado no trabalho de Wilson, Summers e Hope (2000);
- As variáveis testadas neste trabalho são principalmente da criatividade e sensibilidade do autor. Durante a finalização deste estudo, surgiram novos insights sobre variáveis que poderiam melhorar o poder de discriminação do modelo de inadimplência. Entre outras variáveis possíveis de serem testadas estão: (1) clientes que possuem frota própria ou agregada, (2) tamanho da frota (número de caminhões), (3) número de funcionários, (4) idade do cadastro do cliente, (5) número de atrasos em faturas dividido pela idade do cadastro;
- Neste trabalho não foram testadas interações entre as variáveis, exemplo: clientes que trabalham com container e estão na condição de fiadores, entre outros;
- É de se esperar que quanto maiores forem os percentuais de descontos concedidos nas vendas, maiores serão as quantidades compradas e vice-versa. Assim seria interessante incluir no modelo de CLV um mecanismo que corrija a quantidade de produtos vendidos em função do desconto praticado para cada cliente. Outra sugestão seria

adicionar outro mecanismo que corrija a quantidade vendida em função do valor do preço de venda do litro de diesel (elasticidade);

- Para o modelo de inadimplência, foram testadas duas diferentes variáveis (não foram citadas anteriormente). “FinalMês” e “InícioMês”; estas variáveis testaram, através da utilização da técnica de variáveis dummy, se faturas com data de vencimento nos últimos 5 dias úteis do mês e nos primeiros 5 dias úteis do mês são mais inadimplentes que o restante dos dias no mês. A motivação da inclusão desta variável é que as empresas possuem uma série de pagamentos no final ou início do mês (salários, financiamentos, entre outros) e isso poderia dificultar o pagamento de outras despesas no mesmo período. Os testes foram insatisfatórios, pois as duas variáveis se mostraram insignificantes estatisticamente. Porém a ideia de testar períodos de incidência de inadimplência é interessante. Uma sugestão é verificar meses, dias da semana, períodos de safra, entre outros;
- No modelo proposto de CLV, o cálculo do risco de inadimplência desconsidera a aversão ao risco. Uma melhoria no modelo seria incluir um índice de correção do risco de inadimplência relacionado à aversão ao risco da empresa estudada. Um desafio seria encontrar uma forma de incluir este índice no modelo. Além disso, existe a possibilidade de utilizar o ponto de corte do modelo de risco de inadimplência como um índice de aversão ao risco;
- Rejeitar clientes porque o CLV é negativo sem avaliar outras questões importantes pode ser arriscado. Existem clientes que, mesmo apresentando um CLV negativo, são estratégicos para a empresa. Utilizar diferentes pressupostos no modelo pode ajudar a tornar os resultados mais reais. Um exemplo disso é o trabalho de Helm (2004). No estudo, a autora demonstra as consequências de integrar componentes intangíveis, tais como: a influência do cliente no mercado, a propaganda boca a boca, as recomendações, etc;
- Embora esta pesquisa tenha utilizado uma extensa base de dados, o período de análise é relativamente curto (menos de 2 anos). Deve-se melhor estudar a questão do período de análise, pois existem questões de sazonalidade em algumas variáveis estudadas;
- Deve-se estudar com cuidado a questão de utilizar o passado para prever o futuro.

Por fim, no uso gerencial, é importante destacar que o modelo não foi elaborado com a pretensão de tomar decisões para os gestores, mas sim de ajudar nas decisões sobre a base de

clientes. Deve-se também ter cuidado com as classificações dos riscos geradas pelo modelo de risco de inadimplência, já que a definição de risco utilizada é mais rigorosa quando comparada às outras encontradas na literatura.

REFERÊNCIAS BIBLIOGRÁFICAS

AERON H; BHASKAR, T; SUNDARARAJAN, R; KUMAR, A; MOORTHY, J. A Metric for Customer Lifetime Value of Credit Card Customers. **Database Marketing & Customer Strategy Management**, vol. 15, 3, p. 153-168, 2008.

ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Failure. **Journal of Finance**, n.23, p.589-609, 1968.

ALTMAN, E. I; BADYA, T. K. N; DIAS, L. M. R. Previsão de problemas financeiros em empresas. **Revista de Administração de Empresas**, n.19, vol. 1, jan./mar. 1979.

ANDERSON, J. A. Logistic Discrimination. **Handbook of Statistics**, Vol. 2, North Holland Publishing Company, 1982. P. 169-91.

ARAÚJO, E. A; CARMONA, C. U. de M. **Gestão da Inadimplência de uma instituição de Microcrédito com uso de Modelo Credit Scoring**. In: SBFIN, 2008.

ATKIN, B; POCOCK, M. A; TAYLOR, A. H. (Ed.). **Financial planning and control**. 2. Ed. Gower: Pricing Policy, 1988.

BACK, Barbro *et al.* Choosing bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms. **Technical Report**, Turku Centre for Computer Science, Finlândia, n. 40, set. 1996. Disponível em: <www.defaultrisk.com/pp_score_02.htm>. Acesso em dez. 2008.

Banco Central do Brasil (BACEN). Disponível em: <<http://www.bcb.gov.br/?COPOMJUROS>>. Acesso em: 01 de jul. 2010.

BARTH, N. L. **Inadimplência: Construção de Modelos de Previsão**. São Paulo: Nobel, 2004.

BEAVER, W. H. Financial Ratios as Predictors of Failure. **Empirical Research in Accounting: Selected Studies**, suplemento do **Journal of Accounting Research**, autumn, 1966.

BELL, D; DEIGHTON, J; REINARTZ, W. J; RUST, R. T; SWARTZ, G. Seven Barriers to Customer Equity Management. **Journal of Service Research**, v. 5, n. 1: 77-85, 2002.

BERGER, P. D; ECHAMBADI, N; GEORGE, M; LEHMANN, D. R; RIZLEY, R; VENKATESAN, R. From Customer Lifetime Value to Shareholder Value. **Journal of Service Research**, 9(2), 156-167, 2006.

BERGER, P. D; NASR, N. I. Customer lifetime value: Marketing models and applications. **Journal of Interactive Marketing**, v. 12, n. 1 p. 18-30, 1998.

BERTUCCI, L. A. *et al.* Condicionantes de adimplência em processos de concessão de crédito a micro e pequenas empresas. In: III ENCONTRO BRASILEIRO DE FINANÇAS.

2003, São Paulo. **Anais...** São Paulo: Faculdade de Economia, Administração e Contabilidade da USP, 2003.

BRAGA, A. C. S. **Curvas ROC: Aspectos Funcionais e Aplicações**. Portugal, 2000. Tese (Doutorado em Engenharia de Produção e Sistemas) Universidade do Minho.

CAMARGO, L. F. R. **Proposição de um Modelo Baseado em Customer Lifetime Value para a Análise de Melhorias no Sistema Produtivo**. São Leopoldo, 2009. Dissertação (Mestrado Engenharia da Produção) Unisinos.

CAOUCETTE, J. B; ALTMAN, E. I; NARAYANAN, P. **Gestão do Risco de Crédito**. Rio de Janeiro: Ed. Qualitymark, 1999.

DANCEY, C. P; REIDY, J. **Estatística sem Matemática para Psicologia**. Porto Alegre: Artmed , 2006.

DIAS, M. A. P. **Administração de Materiais: uma abordagem logística**. São Paulo: Atlas. 1993.

DWYER, F. R. Customer Lifetime Valuation to Support Marketing Decision Making. **Journal of Direct Marketing**, vol. 3, p. 8-15, 1999.

EIFERT, D. S; **Análise Quantitativa na Concessão de Crédito Versus Inadimplência: Um Estudo Empírico**. Porto Alegre, 2003. Dissertação (Mestrado em Administração) UFRGS.

ESTADÃO.COM.BR. Disponível em: <<http://www.estadao.com.br/noticias/economia,fed-decide-manter-taxa-de-juros-nos-eua-entre-zero-e-025,362603,0.htm>>. Acesso em: 01 de jul. 2010.

FADER, P. S; HARDIE, B. G. S; JERATH, K. Estimating CLV Using Aggregated Data: The Tuscan Lifestyles Case Revisited. **Journal of Interactive Marketing**. v 21, n 3, 2007.

FADER, P. S; HARDIE, B. G. S; LEE, L. K. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model. **Marketing Science**. vol. 24, n. 2, p 275-284, 2005.

FAWCETT, T. An Introduction to ROC Analysis. **Pattern Recognition Letters**, p.861-874. 2005.

GITMAN, L. J. **Princípios de Administração Financeira**. (10ª Ed.). (Sanvicente, A. Z; Trad.). São Paulo: Pearson, 2008.

GUPTA, S; LEHMANN, D. Customer as Assets. **Journal of Interactive Marketing**, vol. 17. n. 1: 9-24, 2003

GUPTA, S; ZEITHAML, V. Customer Metrics and Their Impact on Financial Performance. **Marketing Science**, vol. 25, n. 6, p. 718-739, 2006.

HAIR, J. F; ANDERSON, R. E; TATHAM, R. L; BLACK, W. C. **Análise Multivariada de Dados**. Porto Alegre: Bookman, 2005a.

HAIR, J. F; BABIN, B; MONEY, A. H; SAMOUEL, P. **Fundamentos de Métodos de Pesquisa em Administração**. Porto Alegre: Bookman, 2005b.

- HAUSER, J. R. How Puritan-Bennett used the house of quality. **Sloan Management Review**, (Spring), p. 61–70, 1993.
- HAYES, R; PISANO, G; UPTON, D; WHEELWRIGHT S. **Produção, Estratégia e Tecnologia em Busca da Vantagem Competitiva**. Porto Alegre: Artmed, 2008. 384 p.
- HELM, S. Monetizing the Intangible Values of Customers: The Example of Referrals. **AMA Winter Marketing Educators Conference**. Chicago, USA, 2004.
- HO, R. **Hand Book of Univariate and Multivariate Data Analysis and Interpretation With SPSS**. New York: Chapman & Hall/CRC, 2006. 403 p.
- HOSMER, D. W; LEMESHOW, S. **Applied Logistic Regression**. New York: John Wiley & Sons, 1989. 397 p.
- JAIN, D; SINGH, S. Customer Lifetime Value Research in Marketing: A Review and Future Directions. **Journal of Interactive Marketing**, vol. 16(2), 34-46, 2002.
- KANITZ, S. C. Como Prever Falências de Empresas. **Revista Exame**, p. 95-102, dez. 1974.
- KUMAR, V; VENKATESAN, R; BOHLING, T; BECKMANN, D. The Power of CLV: Managing Customer Lifetime Value at IBM. **Marketing Science**, vol. 27, n. 4, p. 585-599, 2008.
- LEHMANN, D. R. Metrics for Marketing Matter, **Journal of Marketing**, vol. 68, n. 4:73-75, 2004.
- LITTLE, D. C. J. Models and Managers: The Concept of a Decision Calculus. **Management Science**, 50 (12), 1841-1853.
- MALHOTRA, Naresh K. **Pesquisa de Marketing: uma Orientação Aplicada**. 4.ed. Porto Alegre: Artmed, 2006.
- MAULIK, D; YARLAGADDA, P; YOUNGBLOOD, J. P; CISTON, P. **Comparative Efficacy of Umbilical Arterial Doppler Indices for Predicting Adverse Perinatal Outcome**. Am J Obstet Gynecol, 164, 1434-1440. 1991.
- MELLO, G. R; SLOMSKI, V. (2006) Verificando os Endividamentos dos Estados Brasileiros: Uma Proposta Utilizando Análise Multivariada de Dados. **Anais do XXX EnANPAD**. Salvador, BA, Brasil.
- PAIVA, J. C. N. (2004) A Estrutura de Valor para o Cliente Pessoa Física no Varejo Brasileiro: Uma Escala de Percepção de Valor. **Anais do XXVIII EnANPAD**. Curitiba, PR, Brasil.
- RAJU, J. S; ZHANG, Z. J. **Escolher a estratégia de preços errada pode custar caro**. Disponível em: <<http://www.wharton.universia.net/index.cfm?fa=printArticle&ID=600&language=Portugues>>. Acesso em: 17 ago. 2009.
- RIBEIRO, C. F. **Proposta de Construção de Modelo Econométrico para Estimar a Probabilidade de Risco de Inadimplência: Uma Verificação Empírica na Universidade**

Católica de Pelotas. São Leopoldo, 2008. Dissertação (Mestrado em Contabilidade) Unisinos.

ROSS, S. A; WESTERFIELD, R. W; JAFFE, F. J. **Administração Financeira Corporate Finance.** (Sanvicente, A. Z; Trad.). São Paulo: Atlas, 2008.

RUST, R. T; LEMON, K. N; ZEITHAML, V. A Driving Customer Equity: Linking Customer Lifetime Value to Strategic Marketing Decisions. **Marketing Science Institute**, n. 1: 108, 2001.

RUST, R. T; LEMON, K. N; ZEITHAML, V. return on Marketing: Using Customer Equity to Focus Marketing Strategy. **Journal of Marketing**, vol. 68, p.109-127, 2004.

RUST, R. T; ZEITHAML, V; LEMON, K. N. **O Valor do Cliente: O Modelo que está Reformulando a Estratégia Corporativa.** Porto Alegre: Bookman, 2001.

SANVICENTE, A. Z; MINARDI, A. M. A. F. Identificação de Indicadores Contábeis Significativos para Previsão de concordata de Empresas. **FinanceLab Working Paper – FLWP**, n. 3, 1998. Disponível em: <http://www.risktech.com.br/PDFs/indicadores_concordata.pdf>. Acessado em 12 de outubro de 2009.

Serviço de Apoio às Micro e Pequenas Empresas. (SEBRAE). Disponível em: <<http://www.sebrae.com.br/customizado/estudos-e-pesquisas/estudos-e-pesquisas/sobrevivencia-das-micro-e-pequenas-empresas>>. Acesso em: 01 out. 2009.

SCHIMITTLEIN, D. C; MORRISON, D. G; COLOMBO, R. Counting Your Customers: Who Are They and What Will They Do Next? **Management Science**. 33, 1, 1987.

SLACK, N. *et al.* **Administração da Produção.** São Paulo: Atlas. 1997.

SPACKMAN, K. A. Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning. **Proc. Sixth Internat. Workshop on Machine Learning**. p.160-163. 1989.

SWETS, J. A. Measuring the Accuracy of Diagnostic Systems. **Science**. 240, 1285-1293, 1988.

TUBINO, D. F. **Manual de Planejamento e Controle da Produção.** São Paulo: Atlas. 2000.

VENKATESAN, R; KUMAR, V. A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. **Journal of Marketing**, vol. 68, p. 106-125, 2004.

VILLANUEVA, J; HANSSENS, D. M. Customer Equity: Measurement, Management and Research Opportunities. **Foundations and Trends in Marketing**, 1(1), 1-95, 2007.

WILSON, N; SUMMERS, B; HOPE, R. Using Payment Behaviour Data for Credit Risk Modeling. *International Journal of the Economics of Business*, vol. 7, n.3, p. 333-346, 2000.

YIN, R. K. **Estudo de Caso: Planejamento e Métodos.** Porto Alegre: Bookman, 2001.

APÊNDICE A. Série mensal de compra de litros de diesel por cliente.

Mês	589	145	594	245	144	605e844	49e1568	Outros
Jan/08	13.113	30.204	49.496	32.268	151.222	138.120	75.413	466.371
Fev/08	26.620	34.108	77.026	30.271	100.405	138.445	122.163	497.442
Mar/08	81.397	38.103	109.942	51.482	79.424	140.304	100.495	580.672
Abr/08	269.006	50.458	67.020	231.466	100.932	90.837	196.829	1.171.640
Mai/08	113.022	44.314	60.193	192.582	88.923	121.611	238.763	1.192.172
Jun/08	71.514	56.312	76.869	112.791	99.351	171.699	229.494	1.022.300
Jul/08	58.817	77.079	70.268	77.869	92.009	120.480	213.182	946.865
Ago/08	42.653	46.485	67.671	31.540	94.336	121.144	154.910	688.594
Set/08	24.331	63.101	86.121	32.150	138.860	131.013	88.923	721.231
Out/08	23.092	60.229	60.147	35.675	125.212	145.059	110.051	587.557
Nov/08	24.451	44.075	85.284	19.027	104.414	168.003	73.985	410.316
Dez/08	16.128	62.767	113.543	36.072	139.763	186.128	46.454	399.524
Jan/09	29.236	37.802	22.848	51.226	84.090	133.878	181.659	431.730
Fev/09	12.910	68.258	29.771	40.595	77.795	174.304	55.743	367.103
Mar/09	149	88.408	58.910	47.057	85.996	96.286	42.300	583.766
Abr/09	69.678	46.162	49.483	95.123	104.265	-	381.247	737.000
Mai/09	21.965	31.747	56.492	138.110	14.973	-	369.764	583.759
Jun/09	4.417	25.113	76.698	101.814	27.308	-	220.554	652.799
Jul/09	19.160	20.041	63.868	60.219	46.594	-	185.439	516.797
Ago/09	553	18.415	59.203	48.390	43.784	-	54.962	438.540
Média	46.111	47.159	67.043	73.286	89.983	138.487	157.116	649.809

Tabela 22. Série temporal de venda de litros de diesel para os principais clientes da empresa.

APÊNDICE B. Estatística Descritiva.

A seguir, os resultados de algumas das estatísticas descritivas realizadas. O objetivo de se realizar este tipo de estudo é sintetizar a base de dados, permitindo melhor compreensão global dos dados.

Cliente	Número de faturas			Em reais		
	Q total	Q ad	Q inad	R\$ total	R\$ ad	R\$ inad
49e1568	752	710	42	18.781.462,36	18.493.125,31	288.337,05
605e844	561	355	206	12.375.673,67	9.051.915,41	3.323.758,26
144	672	492	180	10.545.887,75	7.296.757,50	3.249.130,25
245	642	622	20	8.535.697,71	8.278.441,22	257.256,49
594	334	321	13	7.831.258,55	7.610.897,75	220.360,80
145	201	-	201	5.438.163,73	-	5.438.163,73
589	176	32	144	5.286.093,62	384.726,06	4.901.367,56
Outros	10.921	9.164	1.757	760.42.411,70	61.256.533,77	14.785.877,93

Quadro 8. Estatística descritiva.

No Quadro 8, estão os sete principais clientes da empresa estudada em valor de faturamento. O Quadro 8 dá uma ideia melhor sobre a ocorrência de inadimplência (em quantidade de faturas e em reais).

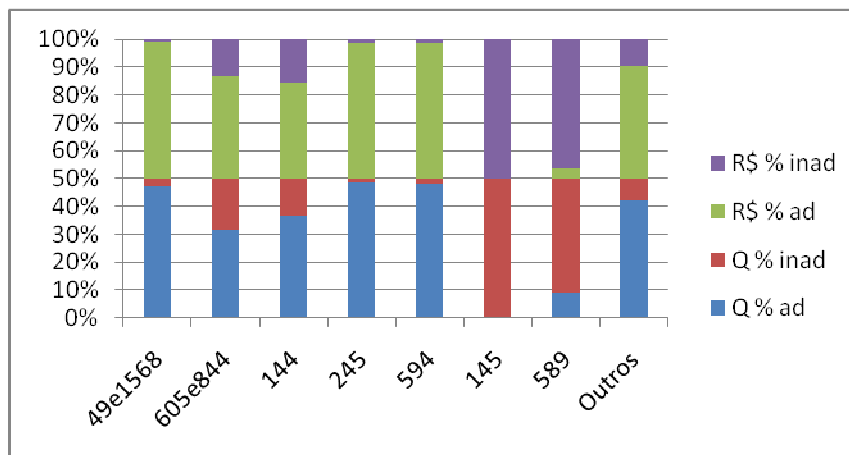


Figura 15. Estatística descritiva.

A Figura 15 é uma representação do Quadro 8. Aqui a medida é em percentuais. As barras em azul e vermelho representam o percentual, em quantidade de faturas, adimplentes e inadimplentes, respectivamente. As barras em roxo e verde representam o percentual, em

reais, da ocorrência de adimplência e inadimplência, respectivamente, nos principais clientes da empresa estudada. Atenção ao cliente “145”, que apresentou a ocorrência de inadimplência em todas as faturas.

Variável	Número de faturas			Em reais		
	Q total	Q ad	Q inad	R\$ total	R\$ ad	R\$ inad
Geral	14.259	11.696	2.563	144.863.649,10	112.399.397,03	32.464.252,07
PercepdaEmp	1.505	847	658	22.376.802,64	11.568.675,48	10.808.127,16
Pagt-1(inad)	2.525	668	1.857	32.656.449,45	6.494.841,89	26.161.607,56
Pagt-2(inad)	2.489	796	1.693	32.380.842,18	7.818.401,95	24.562.440,23
Pagt-3(inad)	2.454	819	1.635	32.101.770,59	8.333.004,52	23.768.766,07
Pelotas+contêiner	267	9	258	5.989.210,25	30.599,13	5.958.611,12
Riogrande+contêiner	1.433	1.065	368	25.138.051,84	20.739.555,09	4.398.496,75

Quadro 9. Estatística descritiva.

No Quadro 9, a linha “Geral” representa toda a base de dados (com dados faltantes). Neste quadro também estão algumas variáveis independentes testadas no modelo de previsão de inadimplência. As ocorrências são descritas de duas diferentes formas: em número de faturas e em reais.

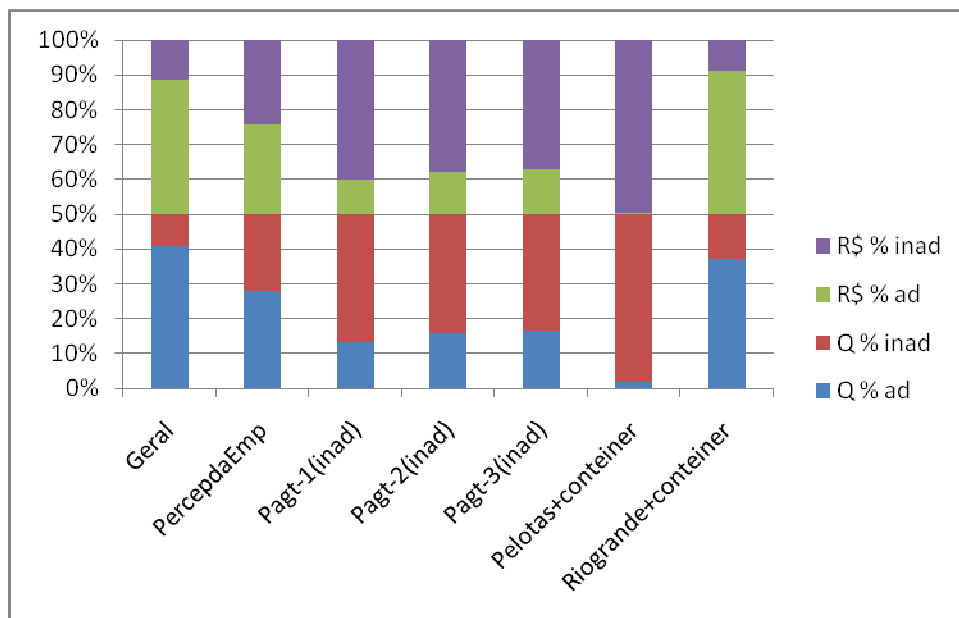


Figura 16. Estatística descritiva.

A Figura 16 é uma representação do Quadro 9. Aqui a medida é em percentuais. As barras em azul e vermelho representam o percentual, em quantidade de faturas, adimplentes e inadimplentes, respectivamente. As barras em roxo e verde representam o percentual, em

reais, da ocorrência de adimplência e inadimplência, respectivamente, nos principais clientes da empresa estudada. Na Figura 16, é observado que os clientes localizados em Pelotas e que trabalham com o tipo de carga de contêiner apresentam maior número de inadimplência.

Cidade	Número de faturas			Em reais		
	Q total	Q ad	Q inad	R\$ total	R\$ ad	R\$ inad
Rio Grande	3.836	3.339	497	60.505.108,34	54.967.645,06	5.537.463,28
Pelotas	3.239	2.495	744	33.978.453,39	22.047.674,91	11.930.778,48
Tupandi	672	492	180	10.545.887,75	7.296.757,50	3.249.130,25
Ijuí	296	152	144	5.526.029,44	624.661,88	4.901.367,56
Porto Alegre	511	366	145	4.113.736,41	3.110.233,43	1.003.502,98
Venâncio Aires	358	173	185	2.774.149,36	1.065.192,08	1.708.957,28
Passo Fundo	367	305	62	2.660.605,22	2.114.838,85	545.766,37
Erechim	150	148	2	1.869.005,9	1.848.552,53	20.453,37
Cruz Alta	237	136	101	1.734.883,78	878.981,49	855.902,29
Santa Cruz do Sul	345	313	32	1.484.307,11	1.210.160,07	274.147,04
Caxias do Sul	190	168	22	1.367.392,71	1.251.216,35	116.176,36
Canoas	386	299	87	1.313.609,43	880.118,77	433.490,66
Carazinho	185	152	33	1.220.616,91	998.002,82	222.614,09
Bento Gonçalves	146	144	2	1.144.822,81	1.140.214,94	4.607,87
Eldorado do Sul	187	77	110	976.299,01	344.887,64	631.411,37
Outros	3.154	144	217	13.621.741,52	12.593.258,7	1.028.482,82

Quadro 10. Estatística descritiva.

No Quadro 10, estão as principais cidades onde os clientes da empresa estudada estão localizados. As ocorrências são descritas de duas diferentes formas: em número de faturas e em reais.

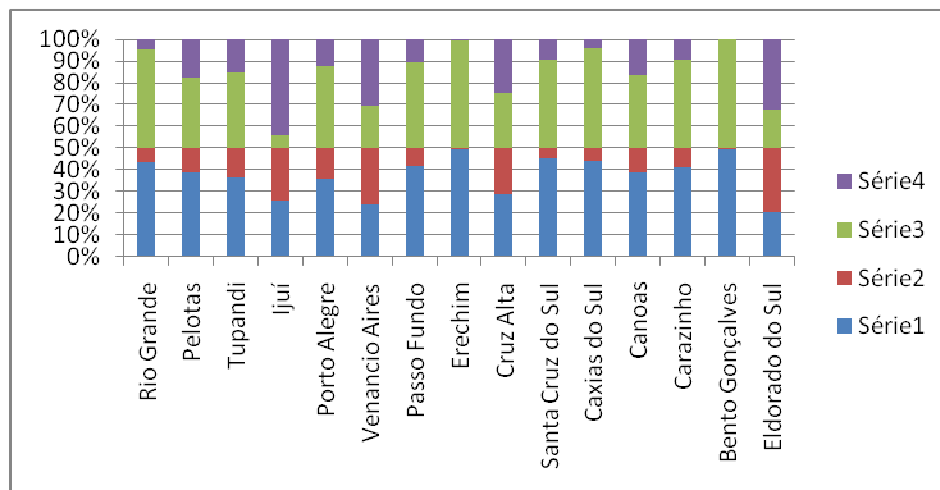


Figura 17. Estatística descritiva.

A Figura 17 é uma representação do Quadro 10. Aqui a medida é em percentuais. As barras em azul e vermelho representam o percentual, em quantidade de faturas, adimplentes e inadimplentes, respectivamente. As barras em roxo e verde representam o percentual, em reais, da ocorrência de inadimplência e inadimplência, respectivamente.

Estado	Número de faturas			Em reais		
	Q total	Q ad	Q inad	R\$ total	R\$ ad	R\$ inad
MG	3	3	-	3.430,00	3.430,00	-
MS	39	39	-	52.754,04	52.754,04	-
MT	18	17	1	37.924,07	34.707,91	3.216,16
PR	346	316	30	1.386.169,91	1.342.799,31	43.370,60
RJ	33	33	-	114.083,00	114.083,00	-
RS	13.698	11.167	2.531	142.551.764,30	110.177.083,79	32.374.680,51
SC	69	68	1	548.342,40	505.357,60	42.984,80
SP	53	53	-	142.181,35	142.181,35	-

Quadro 11. Estatística descritiva.

O Quadro 11 mostra a ocorrência de inadimplência/adimplência de acordo com os estados. As ocorrências são descritas de duas diferentes formas: em número de faturas e em reais.

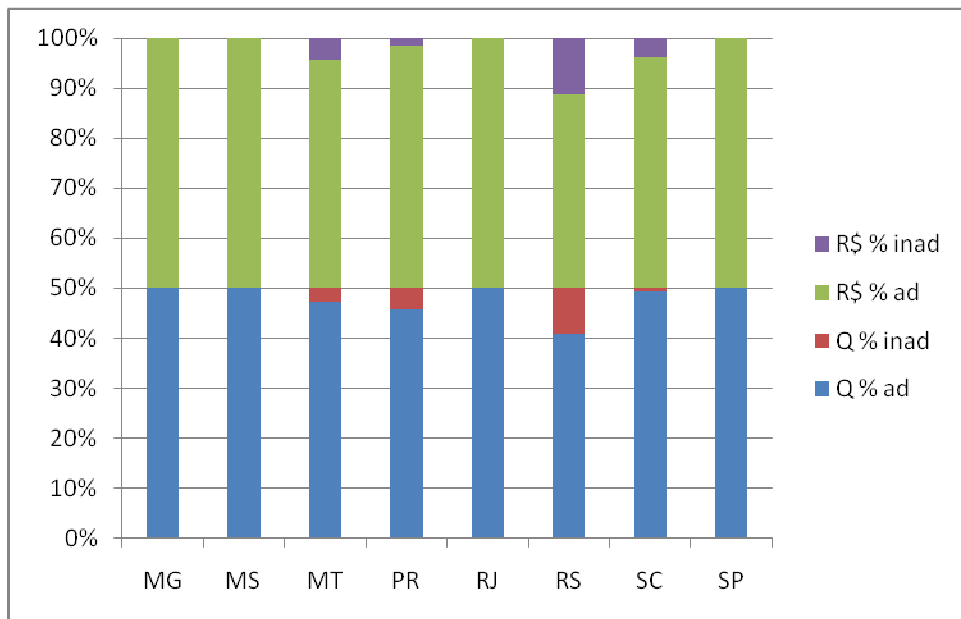


Figura 18. Estatística descritiva.

A Figura 18 é uma representação do Quadro 11. Aqui a medida é em percentuais. As barras em azul e vermelho representam o percentual, em quantidade de faturas, adimplentes e inadimplentes, respectivamente. As barras em roxo e verde representam o percentual, em reais, da ocorrência de adimplência e inadimplência, respectivamente.

Carga	Número de faturas			Em reais		
	Q total	Q ad	Q inad	R\$ total	R\$ ad	R\$ inad
Açúcar	32	31	1	18.302,96	18.101,16	201,80
Adubo	11	11	-	12.410,68	12.410,68	-
Água	120	120	-	239.935,82	239.935,82	-
Areia	12	1	11	42.280,00	6.414,00	35.866,00
Bobina de Papel	5	5	-	9.519,40	9.519,40	-
Câmara Fria	5	5	-	9.685,16	9.685,16	-
Cimento	153	150	3	1.516.036,54	1.514.022,26	2.014,28
Container	3.700	2.633	1.067	49.807.313,20	34.699.675,38	15.107.637,82
Exportação	213	185	28	1.048.868,48	893.211,43	155.657,05
Farinha	7	6	1	3.729,27	3.036,50	692,77
Ferro	261	241	20	788.197,40	763.444,57	24.752,83
Fertilizante	371	267	104	1.764.983,41	1.103.569,63	661.413,78
Fumo	14	14	-	20.915,25	20.915,25	-
Geral	125	123	2	428.569,62	419.023,35	9.546,27
Grãos	8.620	7.392	1.228	86.679.482,96	70.821.483,01	15.857.999,95
Lubrificante	54	50	4	158.569,00	151.841,20	6.727,80
Óleo	43	42	1	24.139,10	23.639,10	500,00
Polietileno	331	319	12	1.092.500,70	1.014.888,30	77.612,40
Telha	33	33	-	75.544,17	75.544,17	-
Trator	149	68	81	1.095.665,97	572.036,65	523.629,32

Quadro 12. Estatística descritiva.

O Quadro 12 mostra a ocorrência de inadimplência/adimplência de acordo com os tipos de cargas transportadas pelos clientes. As ocorrências são descritas de duas diferentes formas: em número de faturas e em reais.

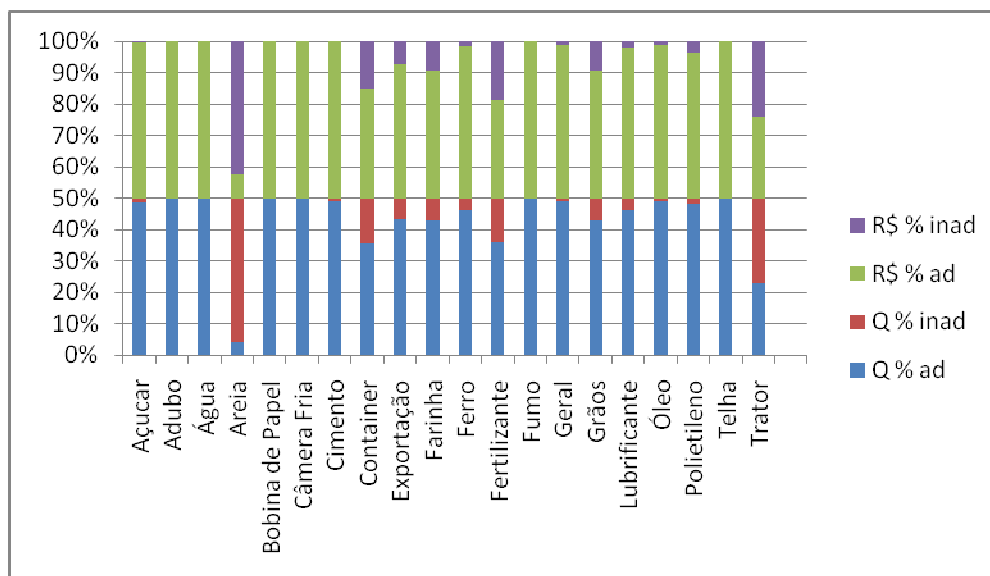


Figura 19. Estatística descritiva.

A Figura 19 é uma representação do Quadro 12. Aqui a medida é em percentuais. As barras em azul e vermelho representam o percentual, em quantidade de faturas, adimplentes e inadimplentes, respectivamente. As barras em roxo e verde representam o percentual, em reais, da ocorrência de adimplência e inadimplência, respectivamente.

APÊNDICE C. Estatística Descritiva do Risco Estimado dos Clientes Estudados.

Cliente	Número de Faturas	Risco estimado			
		Mínimo	Máximo	Média	Desvio Padrão
49e1568	749	0,02	0,71	0,04	0,07
605e844	555	0,03	0,86	0,30	0,36
589	173	0,02	0,76	0,61	0,27
245	639	0,02	0,76	0,04	0,09
145	198	0,89	0,89	0,89	0,00
144	669	0,05	0,89	0,25	0,27
594	331	0,03	0,82	0,05	0,09

Tabela 23. Estatística descritiva.

A Tabela 23 é um resumo do risco de inadimplência estimado para os principais clientes da empresa estudada. A seguir está uma representação gráfica dos riscos estimados para os clientes.

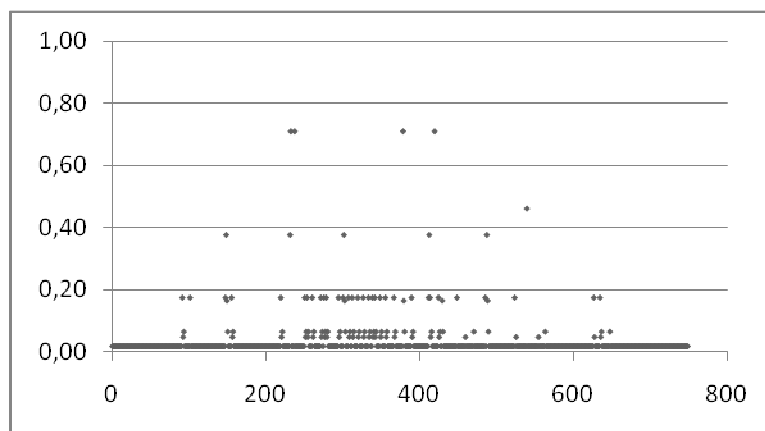


Figura 20. Evolução do risco de inadimplência estimado para o cliente 49e1568.

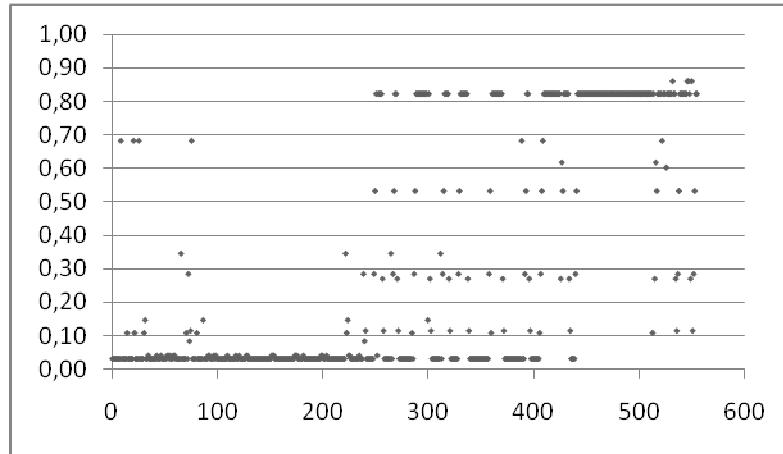


Figura 21. Evolução do risco de inadimplência estimado para o cliente 605e844.

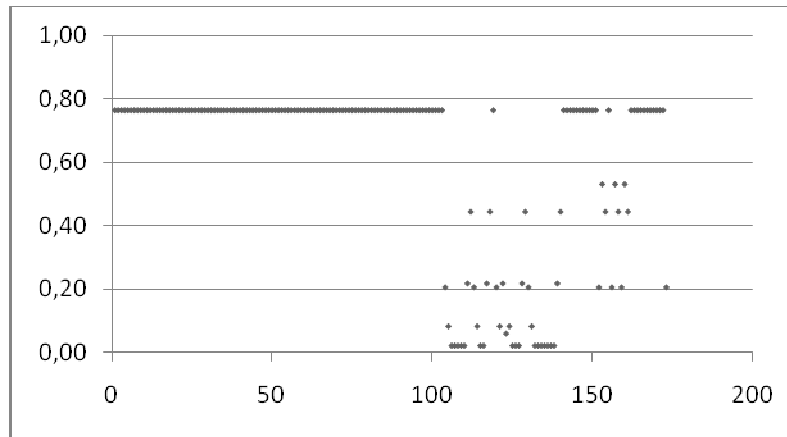


Figura 22. Evolução do risco de inadimplência estimado para o cliente 589.

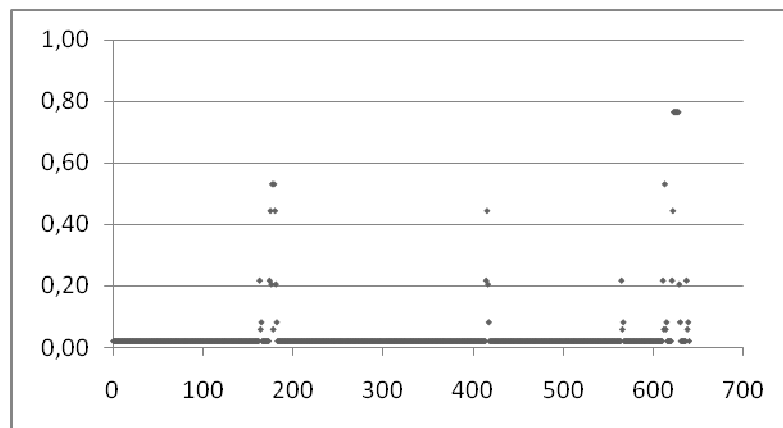


Figura 23. Evolução do risco de inadimplência estimado para o cliente 245.

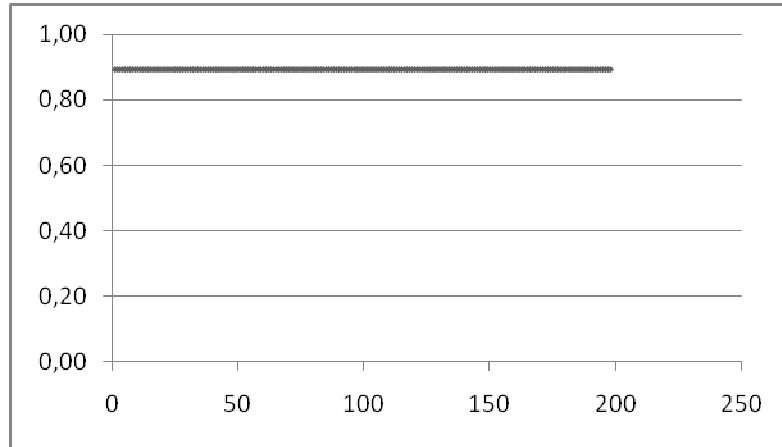


Figura 24. Evolução do risco de inadimplência estimado para o cliente 145.

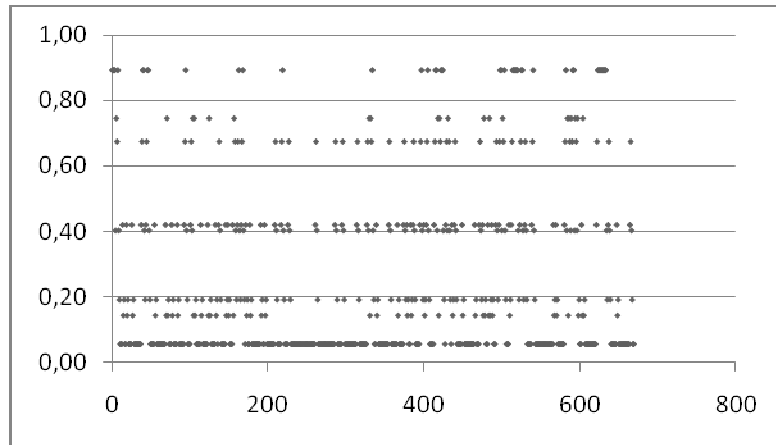


Figura 25. Evolução do risco de inadimplência estimado para o cliente 144.

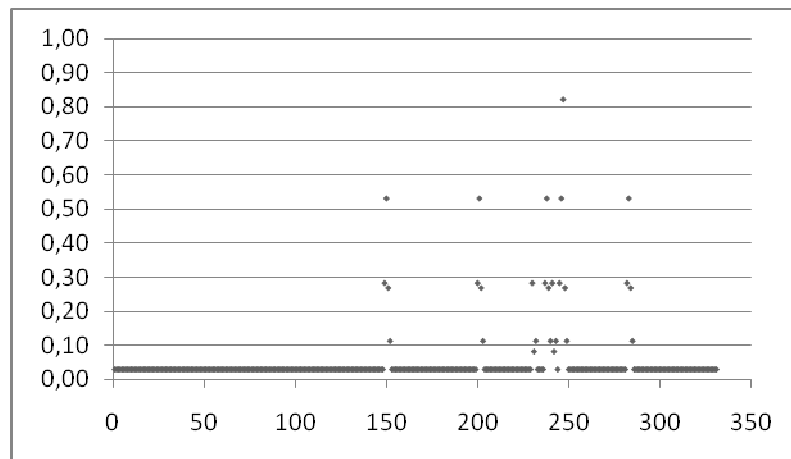


Figura 26. Evolução do risco de inadimplência estimado para o cliente 594.

APÊNDICE D. Resultados curva ROC.

Coordenadas da curva ROC			
Modelo	Ponto de corte	Sensibilidade	1 - especificidade
Completo	0	1	1
	0,02	1,00	1,00
	0,03	1,00	1,00
	0,04	1,00	1,00
	0,05	1,00	1,00
	0,06	1,00	1,00
	0,07	1,00	1,00
	0,08	1,00	0,99
	0,09	1,00	0,99
	0,10	0,99	0,96
	0,11	0,94	0,76
	0,12	0,92	0,68
	0,13	0,84	0,42
	0,14	0,80	0,32
	0,15	0,78	0,29
	0,16	0,75	0,24
	0,17	0,73	0,17
	0,18	0,70	0,12
	0,19	0,69	0,12
	0,20	0,69	0,11
	0,21	0,69	0,11
	0,22	0,68	0,11
	0,23	0,68	0,11
	0,24	0,68	0,11
	0,25	0,68	0,10
	0,26	0,68	0,10
	0,27	0,67	0,10
	0,28	0,67	0,10
	0,29	0,66	0,09
	0,30	0,66	0,09
	0,31	0,65	0,09
	0,32	0,65	0,09
	0,33	0,64	0,08
	0,34	0,63	0,08
	0,35	0,62	0,07
	0,36	0,62	0,07
	0,37	0,62	0,07
	0,38	0,62	0,07
	0,39	0,62	0,07
0,40	0,61	0,07	

0,41	0,61	0,07
0,42	0,60	0,06
0,43	0,60	0,06
0,44	0,60	0,06
0,45	0,60	0,06
0,46	0,59	0,06
0,47	0,59	0,06
0,48	0,57	0,05
0,49	0,57	0,05
0,50	0,57	0,05
0,51	0,57	0,05
0,52	0,56	0,05
0,53	0,55	0,05
0,54	0,54	0,05
0,55	0,54	0,05
0,56	0,54	0,05
0,57	0,53	0,04
0,58	0,53	0,04
0,59	0,52	0,04
0,60	0,52	0,04
0,61	0,52	0,04
0,62	0,51	0,04
0,63	0,50	0,03
0,64	0,49	0,03
0,65	0,49	0,03
0,66	0,49	0,03
0,67	0,49	0,03
0,68	0,48	0,03
0,69	0,48	0,03
0,70	0,48	0,03
0,71	0,48	0,03
0,72	0,48	0,03
0,73	0,47	0,03
0,74	0,47	0,03
0,75	0,46	0,03
0,76	0,45	0,02
0,77	0,44	0,02
0,78	0,44	0,02
0,79	0,42	0,02
0,80	0,41	0,02
0,81	0,40	0,02
0,82	0,40	0,02
0,83	0,40	0,02
0,84	0,40	0,02
0,85	0,39	0,01
0,86	0,39	0,01

	0,87	0,38	0,01
	0,88	0,37	0,01
	0,89	0,37	0,01
	0,90	0,36	0,01
	0,91	0,33	0,01
	0,92	0,26	0,01
	0,93	0,20	0,01
	0,94	0,10	0,00
	0,95	0,06	0,00
	0,96	0,03	0,00
	1,00	0,00	0,00
SemPag	0,00	1,00	1,00
	0,24	0,83	0,76
	0,25	0,68	0,58
	0,26	0,56	0,41
	0,27	0,36	0,17
	0,28	0,25	0,12
	0,29	0,22	0,08
	0,30	0,22	0,07
	0,31	0,22	0,07
	0,32	0,21	0,07
	0,33	0,21	0,07
	0,34	0,21	0,07
	0,35	0,21	0,07
	0,36	0,21	0,07
	0,44	0,21	0,07
	0,52	0,16	0,06
	0,53	0,13	0,06
	0,54	0,13	0,06
	0,55	0,09	0,04
	0,56	0,04	0,01
	0,57	0,02	0,01
	0,58	0,00	0,00
	0,59	0,00	0,00
	0,60	0,00	0,00
	1,00	0,00	0,00
SemVlrRecebido	0,00	1,00	1,00
	0,11	0,95	0,81
	0,14	0,80	0,33
	0,16	0,72	0,15
	0,18	0,70	0,12
	0,20	0,68	0,11
	0,23	0,68	0,11
	0,27	0,67	0,10
	0,30	0,66	0,09
	0,32	0,65	0,09

	0,33	0,63	0,08
	0,35	0,62	0,08
	0,37	0,62	0,07
	0,39	0,61	0,07
	0,40	0,61	0,07
	0,42	0,60	0,06
	0,45	0,60	0,06
	0,46	0,58	0,06
	0,49	0,57	0,05
	0,54	0,54	0,05
	0,57	0,53	0,04
	0,58	0,52	0,04
	0,60	0,51	0,04
	0,62	0,50	0,03
	0,64	0,49	0,03
	0,66	0,49	0,03
	0,68	0,48	0,03
	0,71	0,48	0,03
	0,74	0,45	0,02
	0,76	0,44	0,02
	0,78	0,42	0,02
	0,79	0,41	0,02
	0,80	0,41	0,02
	0,82	0,40	0,02
	0,83	0,40	0,02
	0,84	0,40	0,01
	0,85	0,39	0,01
	0,87	0,38	0,01
	0,89	0,38	0,01
	0,91	0,31	0,01
	0,93	0,11	0,00
	0,95	0,04	0,00
	0,96	0,00	0,00
	1,00	0,00	0,00
SoPag	0,00	1,00	1,00
	0,22	0,68	0,11
	0,33	0,64	0,09
	0,43	0,59	0,06
	0,58	0,52	0,04
	0,71	0,47	0,03
	0,79	0,41	0,02
	0,86	0,38	0,01
	1,00	0,00	0,00

Tabela 24. Pontos de corte para os modelos.