



UNIVERSIDADE SALVADOR
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
MESTRADO EM REGULAÇÃO DA INDÚSTRIA DA ENERGIA

RIBAMAR KLEBER DA SILVA

MODELO DE AVALIAÇÃO DO RISCO DE
INADIMPLÊNCIA DE CONSUMIDORES DE ENERGIA
ELÉTRICA

Salvador
2008

RIBAMAR KLEBER DA SILVA

**MODELO DE AVALIAÇÃO DO RISCO DE INADIMPLÊNCIA DE
CONSUMIDORES DE ENERGIA ELÉTRICA**

Dissertação apresentada ao Curso de Mestrado Profissional em Regulação da Indústria da Energia, Universidade Salvador - UNIFACS, como requisito parcial para obtenção do grau de Mestre.

Orientador: Prof. Dr. José Eduardo Pinheiro Santos Tanure

Salvador
2008

Ficha Catalográfica elaborada pelo Sistema de Bibliotecas da Universidade
Salvador - UNIFACS
FICHA CATALOGRÁFICA

Silva, Ribamar Kleber da
Modelo de avaliação do risco de inadimplência de
consumidores de energia elétrica / Ribamar Kleber da Silva. -
Salvador, 2008.
154 f. : il.

Dissertação (mestrado) - Universidade Salvador - UNIFACS.
Mestrado em Regulação da Indústria de Energia, 2008.

Orientador: Prof. Dr. José Eduardo Pinheiro Santos Tanure.

1. Energia elétrica - Distribuição. 2. Inadimplência (Finanças). 3.
Comportamento do consumidor. I. Tanure, José Eduardo
Pinheiro Santos, orient. II. Título.

CDD: 333.79323

TERMO DE APROVAÇÃO

RIBAMAR KLEBER DA SILVA

MODELO DE AVALIAÇÃO DO RISCO DE INADIMPLÊNCIA DE CONSUMIDORES DE ENERGIA ELÉTRICA

Dissertação aprovada como requisito parcial para a obtenção do grau de Mestre em Regulação da Indústria de Energia, Universidade Salvador – UNIFACS, pela seguinte banca examinadora:

José Eduardo Pinheiro Santos Tanure – Orientador _____
Doutor em Engenharia Elétrica, Universidade de São Paulo (USP).
Universidade Salvador – UNIFACS.

Fernando Monteiro de Figueiredo _____
Doutor em Engenharia Elétrica, Universidade de São Paulo (USP).
Universidade de Brasília – UNB.

Artur Alves Pessoa _____
Doutor em Informática, Pontifícia Universidade Católica, PUC/RJ.
Universidade Federal Fluminense – UFF.

André Luiz de Carvalho Valente _____
Doutor em Engenharia Elétrica, Universidade de São
Paulo (USP).
Universidade Salvador – UNIFACS.

Kleber Freire da Silva _____
Doutor em Engenharia Elétrica, Universidade de São Paulo (USP).
Universidade Salvador – UNIFACS.

Salvador, 19 de setembro de 2008.

Dedico este trabalho a minha querida mãe Dona Carminha, que me ensinou a preservar a ternura após a perda precoce da figura paterna.

A Deda, minha esposa e companheira de todas as horas, pelo estímulo e convicção de que é possível ter firmeza de propósito e conseguir realizar nossos sonhos.

A meus filhos Gustavo, Pedro, Julian e Rodolfo. Através deles aprendi a amar como fui amado pelos meus pais.

AGRADECIMENTOS

A gratidão é imensa. Sou eternamente grato a Deus e a todas as pessoas que contribuíram de alguma forma para a realização deste trabalho.

Sou grato ao meu orientador, Prof. Dr. José Eduardo Tanure, cujos conhecimentos, simplicidade e segurança tornaram este trabalho prazeroso e possível de ser realizado.

À banca examinadora pela leitura, análise e comentários.

À Lucimar Rocha “Luci”, a quem sou especialmente grato pelo estímulo sempre renovado por constantes conversas sobre o andamento da pesquisa.

Ao Professor Renato José Pinho de Araújo por encorajar-me nessa empreitada e pelas importantes sugestões.

Aos colegas da 5ª turma do Mestrado em Regulação da Indústria de Energia, em especial a Hugo Machado, Ivana Maciel, Eva Maria e Laucides Damasceno, companheiros da Coelba.

À equipe da Unidade de Estratégia e Gestão da Cobrança pelo apoio e participação na pesquisa e na implantação de seus resultados para melhoria no processo de gestão da cobrança. Faço questão de citar seus nomes: Patrícia Santa Bárbara, Mírian Faustino, Patrícia Moreau, Reginaldo Ramos, Lindinalva Bacelar, Carlos Augusto Guimarães, Isael Passos, Belarmino Dourado, Marcos Vinício e Raquel Cerqueira.

À Antônio Carlos da Costa Pinhel, por inspirar a busca de métodos quantitativos que auxiliem à tomada de decisão, particularmente na gestão de grandes volumes de operações de cobrança.

A Coelba, pelo suporte financeiro e investimento na minha qualificação profissional, em especial, às equipes dos Departamentos de Grandes Clientes e Departamento de Gestão Comercial, onde trabalhei durante o mestrado e principalmente pelo estímulo na conclusão deste trabalho.

Os eventos presentes estão ligados aos eventos precedentes por um vínculo baseado no princípio óbvio de que uma coisa não pode ocorrer sem uma causa que a produza

[...] Todos os eventos, mesmo aqueles que, devido à insignificância, não parecem seguir as grandes leis da natureza, resultam delas tão necessariamente como as revoluções do sol.

(LAPLACE *apud* BERNSTEIN, 1997, p.198).

RESUMO

A reestruturação do setor elétrico brasileiro, iniciada em 1995, inaugurou um novo modelo de regulação dos serviços públicos de distribuição de energia elétrica. Especificamente para as concessionárias de distribuição, o atual marco regulatório estabelece mecanismos de indução à eficiência econômica do monopólio. Nesse modelo, os ganhos de produtividade auferidos pela concessionária são repassados ao consumidor em intervalos regulatórios. Na direção de alcançar ganhos de produtividade na gestão dos recebíveis e na política de crédito e cobrança, os modelos de avaliação do risco de inadimplência baseados na classificação de consumidores pelo comportamento de pagamento (*behavioural scoring*) e na previsão de inadimplência (*credit scoring*) surgem como importantes ferramentas de apoio à decisão. Esses modelos são aplicados em uma concessionária de distribuição de energia elétrica. Para tanto, dados de uma amostra de consumidores da Coelba foram utilizados para desenvolver os estudos. As técnicas estatísticas utilizadas na construção do modelo foram: regressão logística binária e análise de agrupamentos. Os resultados da análise de regressão logística indicam que o modelo alcançou uma classificação correta de consumidores inadimplentes em torno de 84%. Os resultados da análise de agrupamentos indicam três perfis distintos de consumidores, com base no histórico de pagamento dos últimos seis meses. O perfil BS1 agrupou os casos com 91% dos pagamentos realizados em até oito dias do vencimento, o melhor perfil de pagamento. Os casos com uma situação intermediária foram agrupados no perfil BS2, onde 25% dos pagamentos são efetuados apenas com recebimento da fatura, sem a utilização de meios de cobrança adicionais. Já o terceiro perfil, o BS3, agrupou os casos com pior perfil de pagamento, contendo apenas 6% de pagamento em até oito dias da data do vencimento da fatura e 70% de pagamento realizado após a suspensão do fornecimento. Os estudos apresentados neste trabalho confirmam a contribuição das técnicas de modelagem estatística para redução das despesas com as operações de cobrança e com as perdas financeiras com devedores incobráveis, através do controle do risco de inadimplência da carteira de consumidores.

Palavras-chave: Cobrança. Risco de Inadimplência. Comportamento de Pagamento. Distribuição de Energia Elétrica. Regulação de Serviços Públicos.

ABSTRACT

The restructuring of the Brazilian electric sector, which began in 1995, opened a new model for regulation of public services of electric energy distribution. Specifically to the distribution's concessionaires, the current regulatory landmark establishes induction mechanisms to the economic efficiency of the monopoly. In this model, the profits of productivity gained by the concessionaire are transferred to the customer in regulatory intervals. In the direction of achieving productivity gains in the management of accounts receivable and in credit policy and collection, the models for assessing the risk of default are based on the classification of the consumer behavior of payment (behavioural scoring), and in forecast of default risk (credit scoring) arise as important tools for decision support. These models are applied in a distribution's concessionaire of electric energy. To this end, data from a sample of consumers of Coelba were used to develop the studies. The statistical techniques used in the construction of the model were: binary logistic regression and cluster analysis. The results of the logistic regression analysis indicate that the model reached a correct classification of default consumers around 84%. The results of the cluster analysis indicate 3 profiles of different clusters of consumers, based on the payment history of the past 6 months. The profile BS1 joined the cases with 91% of the payments made in up to 8 days of bill due date, the best profile of payment. The cases with a situation intermediate were joined in the profile BS2, where 25% of payments are made only on receipt of the invoice, without the use of additional resources for recovery of the default debt. Already the third profile, the BS3, joined the cases with worse profile payment, containing only 6% of payment in up to 8 days of the bill due date and 70% of payment made after the suspension of supply. The studies presented in this paper confirm the contribution of statistical modeling techniques to reduce expenditure with the operations of recovery and the financial losses with bad debtors, through control of the risk of default of the portfolio of consumers.

Key-Words: Collection. Risk of Default. Credit Scoring Behavioral Scoring Electric. Energy Distribution Insolvency. Public Services Regulation.

LISTA DE FIGURAS

Figura 1 - Cheques sem fundo: 2ª devolução em cada 1.000 compensados.....	18
Figura 2 - Evolução dos índices de inadimplência (%): Contas Vencidas/Faturamento Acumulado nos últimos 12 meses. Dados com PDD. 15 distribuidoras = 62% do mercado.....	19
Figura 3 - Índice de Inadimplência por Segmento (%): Contas Vencidas/Faturamento Acumulado nos últimos 12 meses Dados com PDD. 15 distribuidoras = 62% do mercado.....	19
Figura 4 - Exemplo de distância euclideana entre dois objetos medidos sobre duas variáveis.....	56
Figura 5 - Regra de tratamento de variáveis dicotômicas (binárias).....	57
Figura 6 - Métodos aglomerativo e divisivo.....	59
Figura 7 - Dendrograma ilustrativo do processo de agrupamento hierárquico.....	60
Figura 8 - Algoritmo ligação individual (single linkage method ou nearest neighbor).	61
Figura 9 - Algoritmo ligação completa (<i>complete linkage method ou farthest neighbor</i>)...	61
Figura 10 - Configuração dos quatro objetos para o agrupamento não-hierárquico ..	65
Figura 11 - Grupos AB e CD com os respectivos centróides.....	65
Figura 12 - Grupos A e BCD com o respectivo centróide.....	66
Figura 13 - Distâncias ao quadrado dos Grupos A e BCD.....	67
Figura 14 - Diagrama de perfil das sete observações com base nas cinco variáveis de agrupamentos.....	68
Figura 15 - Dendrograma do agrupamento hierárquico para as sete observações.....	72
Figura 16 - Forma da relação logística entre variáveis dependente e independente...	75
Figura 17 - Dois exemplos hipotéticos de ajuste de uma relação logística aos dados de uma amostra.....	76
Figura 18 - Perfis de agrupamento para solução de dois agregados.....	100
Figura 19 - Perfis de agrupamento para solução de três agregados.....	101
Figura 20 - Distância de Cook em relação à probabilidade predita no modelo logístico	110

LISTA DE TABELAS

Tabela 1 - Comparativo de indicadores.....	17
Tabela 2 - Inadimplência (%) - Contas Vencidas/Faturamento Acumulado nos últimos 12 meses. (dados com PDD). Amostra de 15 distribuidoras = 62% do mercado.....	40
Tabela 3 - Um exemplo de sistema de classificação de empréstimos: símbolos de <i>rating</i> de créditos seniores de longo prazo.....	43
Tabela 4 - Valores de dados das sete observações com base nas cinco variáveis de agrupamento	68
Tabela 5 - Matriz de proximidades de distâncias euclidianas entre observações	70
Tabela 6 - Processo de agrupamento hierárquico aglomerativo para as sete observações.....	71
Tabela 7 - Solução com dois agrupamentos da análise não-hierárquica com pontos sementes iniciais obtidos dos resultados hierárquicos	72
Tabela 8 - Distribuição do número de consumidores com débito por classe e faixa de consumo	81
Tabela 9 - Distribuição do consumo de consumidores com débito por classe e faixa de consumo	82
Tabela 10 - Segmentos de consumidores privados de energia elétrica.....	84
Tabela 11 - Análise do coeficiente de aglomeração para análise hierárquica de agrupamentos pelo método centróide.....	88
Tabela 12 - Perfis das variáveis com as soluções de dois e seis agregados da análise hierárquica de agrupamentos	89
Tabela 13 - Teste de significância de diferenças entre centros de agrupamento.....	90
Tabela 14 - Distribuição de consumidores por atividade econômica: 2 agrupamentos	91
Tabela 15 - Distribuição de consumidores por atividade econômica: 6 agrupamentos	92
Tabela 16 - Análise do coeficiente de aglomeração para análise hierárquica de agrupamentos pelo método de Ward.....	96
Tabela 17 - Solução de dois agrupamentos e de três agrupamentos da análise não-hierárquica com pontos sementes iniciais obtidos dos resultados hierárquicos.....	97
Tabela 18 - Teste de significância de diferenças entre centros de agrupamento para dois e três agrupamentos.....	98
Tabela 19 - Comportamento de pagamento por tipo de cobrança e segmento de cobrança	99
Tabela 20 - Teste de Hosmer-Lemeshow.....	106

Tabela 21 - Tabela de classificação com os resultados práticos do modelo de regressão logística	107
Tabela 22 - Detalhamento das medidas de exame das variáveis incluídas no modelo de regressão logística	108
Tabela 23 - Variáveis preditoras na partida da regressão logística	111
Tabela 24 - Variáveis preditoras que ficaram fora do modelo de regressão logística	112
Tabela 25 - Histórico de interação do modelo de regressão logística.....	112
Tabela 26 - Resultados das medidas de R ² de Cox e Snell, de Nagelkerke e "pseud R ²	114
Tabela 27 - Resultados do método <i>stepwise</i> inverso no modelo de regressão logística	115

SUMÁRIO

1 INTRODUÇÃO	14
1.1 CONTEXTO DA INADIMPLÊNCIA NA DISTRIBUIÇÃO DE ENERGIA ELÉTRICA NO BRASIL	14
1.2 OBJETIVO DO ESTUDO	20
1.3 ORGANIZAÇÃO DO ESTUDO.....	21
2 ANÁLISE REGULATÓRIA DA INADIMPLÊNCIA	23
2.1 O PAPEL DA DISTRIBUIDORA NA ARRECADAÇÃO DE RECURSOS	23
2.2 VISÃO GERAL DA REGULAÇÃO POR INCENTIVOS.....	24
2.3 REGULAÇÃO DA INADIMPLÊNCIA.....	27
2.4 EFEITOS DA INADIMPLÊNCIA SOBRE A PRODUTIVIDADE.....	28
2.4.1 A base de cálculo da inadimplência regulatória	29
2.4.2 Perdas por inadimplência de consumidores	32
2.4.3 Custos de captação do capital de giro	32
2.4.4 Custos operacionais da cobrança	33
3 ABORDAGEM TEÓRICA SOBRE RISCO DE INADIMPLÊNCIA	36
3.1 RISCO DE INADIMPLÊNCIA COM CONSUMIDORES DE ENERGIA ELÉTRICA	39
3.2 MODELOS DE MEDIÇÃO DO RISCO DE INADIMPLÊNCIA	40
4 TÉCNICAS DE ANÁLISE MULTIVARIADA UTILIZADAS	47
4.1 ANÁLISE MULTIVARIADA DE DADOS: DEFINIÇÕES E CONCEITOS GERAIS	47
4.1.1 A variável estatística	48
4.1.2 Escalas de medida	48
4.1.3 Erro de medida e medida multivariada	49
4.1.4 Significância estatística e Poder estatístico	50
4.2 EXAME DOS DADOS	52
4.3 ANÁLISE DE AGRUPAMENTOS (CLUSTER ANALYSIS)	54
4.3.1 Medidas de similaridade	55
4.3.1.1 Coeficientes binários ou dicotômicos	56
4.3.2 Métodos de agrupamento	58
4.3.2.1 Procedimentos hierárquicos de agrupamento	59
4.3.2.2 Procedimentos não-hierárquicos de agrupamento.....	63
4.3.3 Exemplo de uma análise de agrupamentos	67
4.3.4 Objetivo da análise de agrupamentos	69
4.3.5 Projeto de pesquisa em análise de agrupamentos	69
4.3.6 Determinação de agrupamentos e avaliação do ajuste geral	70
4.3.7 Interpretação e validação do perfil dos agrupamentos	73
4.4 ANÁLISE DE REGRESSÃO LOGÍSTICA	73
4.4.1 Estimação do modelo de Regressão Logística	74
4.4.2 Interpretação dos coeficientes	77
4.4.3 Avaliação da adequação do ajuste do modelo e teste da significância dos coeficientes	78

5 ESTUDO DE CASO	80
5.1 SEGMENTAÇÃO DA CARTEIRA DE CONSUMIDORES.....	80
5.1.1 Objetivos da segmentação e análise dos dados dos consumidores inadimplentes.....	80
5.2 ESTUDO DO COMPORTAMENTO DE PAGAMENTO.....	85
5.2.1 Análise de agrupamentos com base na caracterização da dívida.....	85
5.2.1.1 Análise hierárquica de agrupamentos.....	87
5.2.1.2 Identificação de observações atípicas	88
5.2.1.3 Análise de agrupamentos não-hierárquica.....	89
5.2.2 Análise de agrupamentos com base no histórico de pagamentos.....	93
5.2.2.1 Análise hierárquica de agrupamentos.....	94
5.2.2.2 Identificação de observações atípicas	96
5.2.2.3 Análise de agrupamentos não-hierárquica.....	97
5.3 MEDIÇÃO DO RISCO DE INADIMPLÊNCIA PARA A CARTEIRA DE CONSUMIDORES DA COELBA.....	102
5.3.1 Seleção de variáveis dependente e independentes.....	103
5.3.2 Análise dos dados.....	104
6 CONCLUSÃO	117
6.1 CONSIDERAÇÕES FINAIS.....	117
6.2 RECOMENDAÇÕES	119
6.3 DESENVOLVIMENTO DE TRABALHOS FUTUROS.....	120
REFERÊNCIAS.....	122
ANEXO A – Esquema de aglomeração de análise hierárquica pelo método Centróide.....	129
ANEXO B – Dendrograma de análise hierárquica pelo método centróide	134
ANEXO C – Esquema de aglomeração de análise hierárquica pelo método de Ward.....	144

1 INTRODUÇÃO

1.1 CONTEXTO DA INADIMPLÊNCIA NA DISTRIBUIÇÃO DE ENERGIA ELÉTRICA NO BRASIL

A inadimplência de consumidores de energia elétrica chega a comprometer 6,8% do faturamento das concessionárias distribuidoras de energia elétrica (ABRADEE, 2006)¹. Por outro lado, o modelo atual de regulação do setor elétrico brasileiro, denominado regulação pelo preço (*price-cap*), agrupa as concessionárias em três grupos e fixa valores de referência para as receitas irrecuperáveis para cada grupo. O nível de inadimplência do grupo 1 é 0,90%, enquanto o do grupo 2 é 0,60% e o do grupo 3 é de 0,20% (Nota Técnica n.º. 349/2007-SRE) (ANEEL, 2007). Ainda de acordo com o modelo de regulação vigente, níveis de inadimplência acima dos valores admitidos como eficientes não podem ser cobrados na tarifa do consumidor.

Portanto, não compõem a receita anual permitida da concessionária do serviço público de distribuição de energia elétrica² (ANEEL, 2003a, p.20).

Além dos custos com receitas irrecuperáveis decorrentes da inadimplência dos consumidores, os custos com as operações de combate à inadimplência não possuem cobertura tarifária. Portanto, não são considerados na composição de custos da empresa de referência. Para cobertura desses custos, verifica-se apenas a cobrança de uma taxa de religação, que para um consumidor monofásico da Companhia de Eletricidade do Estado da Bahia (Coelba), é de R\$ 4,27³, valor que segundo relatório da Associação Brasileira de Distribuidoras de Energia Elétrica (ABRADEE) sobre custos de serviços taxados não é compatível com os gastos praticados com os serviços de suspensão de fornecimento e religação, além de outras ações de cobrança que serão apresentadas neste estudo.

¹ Dados atualizados pela Abradee. A versão inicial foi apresentada em 2005 no Seminário Abradee de Melhores Práticas (SAMP) – Gestão operacional: gestão da inadimplência.

² Audiência Pública AP 014/2003: Revisão tarifária periódica da concessionária de distribuição de energia elétrica do Estado da Bahia – COELBA. Nota técnica n.º. 052/2003 – SRE/ANEEL. Brasília, março de 2003.

³ Resolução Homologatória n.º. 638, de 17/04/2008 (ANEEL, 2008).

A diversidade de fatores geográficos, sócio-econômicos, técnicos e regulatórios tornam a ação de controle da inadimplência na distribuição de energia elétrica no Brasil, e em especial na região Nordeste, bastante complexa.

Adicionalmente, as disposições previstas nas condições gerais de fornecimento estabelecida pela ANEEL, através da Resolução 456/2000/ANEEL, são insuficientes para garantir baixos níveis de inadimplência, e com isso, o equilíbrio econômico-financeiro das distribuidoras (ANEEL, 2000).

Para garantir a eficiência da arrecadação exigida no atual contexto do setor elétrico brasileiro, as empresas mais bem sucedidas criaram estruturas específicas voltadas para a gestão dos processos de cobrança. Essas estruturas ampliaram as práticas de cobrança, indo além da emissão de reavisos de vencimento e suspensão de fornecimento (corte). Foram implantadas ações tais como: tele-cobrança, inscrição em serviço de proteção ao crédito, cobrança domiciliar, assessoria de cobrança, emissão e protesto de duplicatas e cobrança judicial. Adicionalmente, a estas práticas tem-se realizado grandes investimentos introduzindo novos padrões tecnológicos nos circuitos de baixa tensão para mitigar as ações de auto-religações.

Tomando a concessionária de distribuição baiana como referência, no ano de 2006, o envio mensal de cartas de cobrança envolveu um quarto dos seus quatro milhões de consumidores, a um custo médio de cinquenta centavos. O corte do fornecimento e a religação totalizaram 870 mil operações no ano, a custo médio de R\$ 10,50. Cobrança domiciliar e inscrição em cadastros de maus pagadores totalizam 2,2 milhões de operações a um custo médio de R\$3,50 e R\$1,00, respectivamente. Outras medidas de cobrança são utilizadas, a exemplo de protesto de títulos e assessoria de cobrança.

No Brasil, verifica-se nos anos recentes um expressivo crescimento nas operações de crédito do sistema financeiro, sobretudo no crédito para pessoa física (CrPF). Entre dezembro de 2004 e dezembro de 2007, a participação do volume de crédito sobre o Produto Interno Bruto (PIB) cresceu de 27,0% para 34,7%⁴. Verifica-se um expressivo aumento no volume de crédito concedido a pessoas físicas, que

⁴ Banco Central do Brasil. Refere-se ao crédito do sistema financeiro, exceto com intermediários financeiros, tendo como base o documento contábil Estatística Econômico-Financeira (Estfin).

alcançou R\$ 314,286 bilhões em dezembro de 2007, valor 2,3 vezes maior frente a dezembro de 2004, que representava R\$ 136,442 bilhões (BANCO CENTRAL, 2008). Apesar de o Brasil possuir elevadas taxas de juros, o CrPF vem apresentando um vertiginoso crescimento e provocando movimentos estruturais em novas tecnologias de sistemas de crédito e cobrança e desenvolvimentos de novos produtos.

Pereira (2006, p. 33), ao analisar os movimentos do mercado de créditos CrPF, relata a disposição do consumidor em aceitar um cenário com taxas de juros entre as maiores do mundo e suas implicações, a exemplo de baixo crescimento econômico e desaceleração do ritmo industrial e desemprego. Portanto, diante de um quadro complexo, em que variáveis como dificuldade financeira, desemprego e aumento das tarifas públicas em maior proporção do que o aumento dos salários verifica-se uma fragilidade perigosa que pode potencializar o aumento da inadimplência nas distribuidoras de energia elétrica (DEE), especificamente naquelas com predominância de consumidores pessoa física (COELBA, 2002).

Uma análise dos dados do Banco Central (BACEN), Instituto Brasileiro de Geografia e Estatística (IBGE) sobre a evolução do CrPF, da renda média real no Brasil e na Bahia e do volume de vendas entre dezembro de 2004 e dezembro de 2006, pode indicar que um período de turbulências está em andamento (PEREIRA, 2006). Na Tabela 1, é apresentada uma notável disparidade entre as variações do CrPF e os demais indicadores. Observa-se que a demanda por CrPF é muito superior ao volume de vendas e, ainda mais preocupante, à renda média dos brasileiros e baianos (PEREIRA, 2006). Portanto, esse cenário corrobora com a necessidade de maiores esforços e iniciativas inovadoras para controle da inadimplência nas DEE's.

Tabela 1 - Comparativo de indicadores

INDICADOR	UNIDADE	DEZ/04	DEZ/05	DEZ/06	DEZ/07	DEZ/04 - DEZ/07
Crédito pessoa física- Brasil	R\$ Milhões	136.442	188.784	235.816	314.286	130,3%
Volume de vendas - Brasil *	% (Var. 12 meses)	9,25%	4,76%	6,16%	9,60%	21,9%
Volume de vendas - Bahia *	% (Var. 12 meses)	8,26%	7,08%	9,73%	10,00%	29,3%
Salário mínimo	R\$	260,00	300,00	350,00	380,00	46,2%
Rendimento médio- Brasil	R\$	909,40	1.012,50	1.087,90	1.164,16	28,0%
Rendimento médio- RMS**	R\$	709,50	808,90	858,30	827,10	16,6%

Fontes: Banco Central do Brasil e IBGE.

Notas: (*) Variação do volume de vendas no comércio varejista por Unidade da Federação.

(**) RMS - Região Metropolitana de Salvador.

Em dezembro de 2007, o indicador Serasa de inadimplência dos consumidores, que contempla os registros de cheques devolvidos, títulos protestados, dívidas vencidas com instituições financeiras e cartões de crédito, registrou uma expressiva alta de 18,4% quando comparado com o mesmo período de 2004. Esse indicador utiliza em sua composição a variação da quantidade de cheques sem fundo devolvidos. De acordo com a Figura 1, o indicador que mede a quantidade de cheques sem fundo na segunda devolução em cada 1.000 compensados, aumentou de 15,8 para 18,7 cheques devolvidos entre dezembro de 2004 e dezembro de 2007.

De acordo com estudos da Serasa (2007), os fatores que mais tem determinado o agravamento da inadimplência tem sido a expressiva evolução do CrPF, o maior comprometimento da renda dos consumidores com os parcelamentos assumidos em prazos mais longos e a concessão de crédito sem a utilização de metodologia adequada para análise de risco de crédito e o gerenciamento do risco de inadimplência.

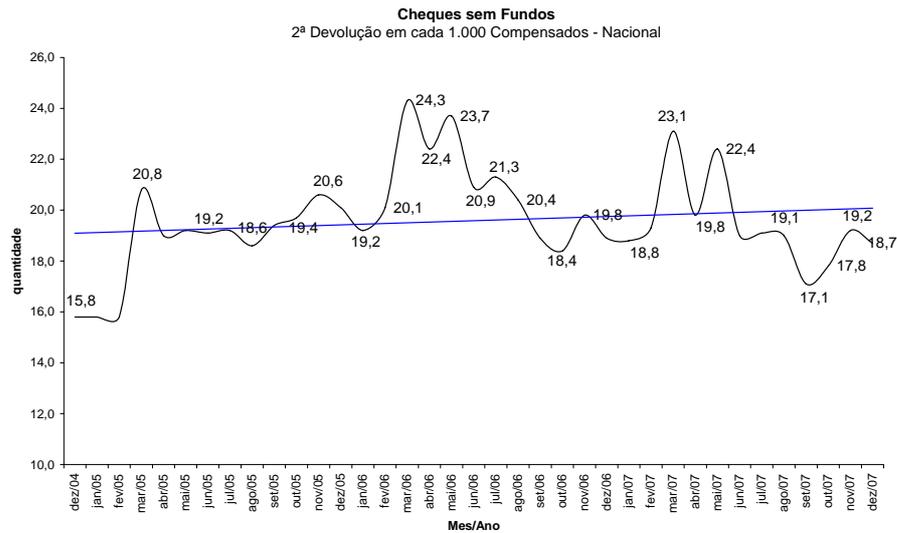


Figura 1 - Cheques sem fundo: 2ª devolução em cada 1.000 compensados

Fonte: Serasa (2008)...

Nota: Adaptação do autor.

Apesar do cenário adverso, o índice que mede a inadimplência total no setor de DEE recuou 1,7% entre junho de 2003 e junho de 2006, ou seja, de 8,5% para 6,8%. A inadimplência até 90 dias também recuou 0,5% no mesmo período. Na Figura 2 é apresentado o índice divulgado pela ABRADDEE e faz referência a quinze concessionárias de distribuição de energia, que representam 62% do mercado brasileiro. As duas métricas de inadimplência apresentadas, denominadas “inadimplência total” e “inadimplência até 90 dias”, são obtidas através da comparação relativa do valor da dívida vencida com o faturamento acumulado dos últimos doze meses.

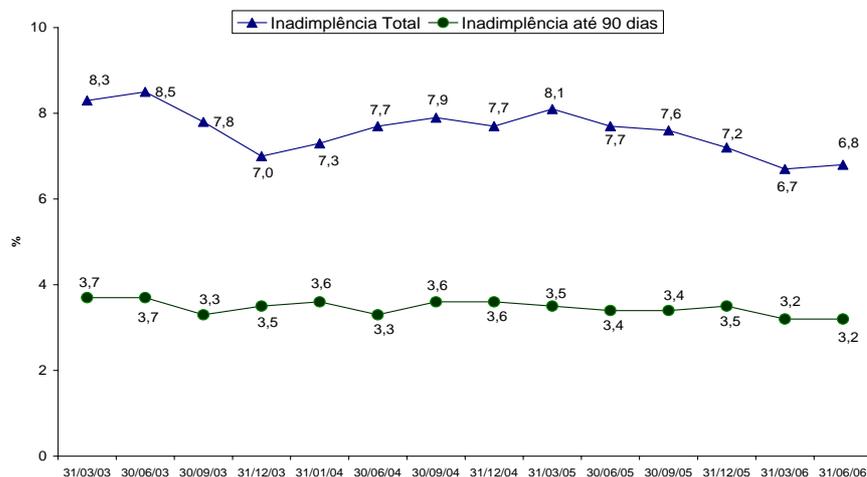


Figura 2 - Evolução dos índices de inadimplência (%): Contas Vencidas/Faturamento Acumulado nos últimos 12 meses. Dados com PDD. 15 distribuidoras = 62% do mercado

Fonte: CVM.

Nota: Elaboração Abradee (2006).

Para a mesma amostra, na análise por segmento observa-se na Figura 3 uma grande distorção na inadimplência do setor público que é três vezes superior à média.

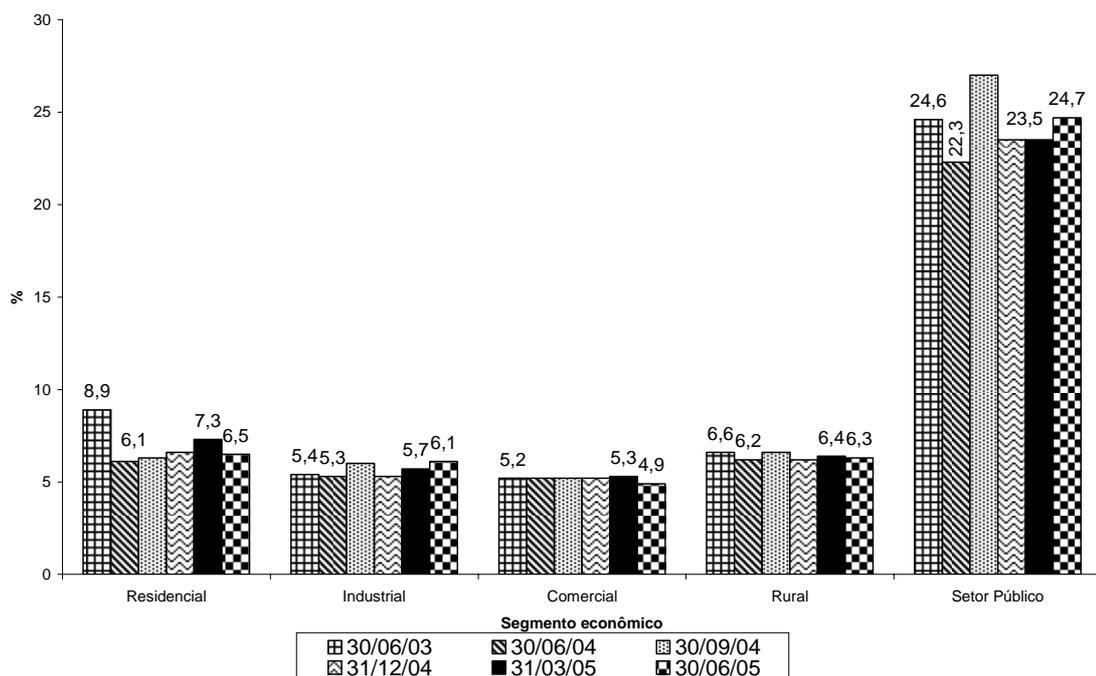


Figura 3 - Índice de Inadimplência por Segmento (%): Contas Vencidas/Faturamento Acumulado nos últimos 12 meses. Dados com PDD. 15 distribuidoras = 62% do mercado

Fonte: CVM.

Nota: Elaboração Abradee (2006).

Do contexto apresentado, pode-se inferir que as DEE, apesar dos esforços de cobrança, estabilizaram a inadimplência em níveis muito superiores aos permitidos no atual arcabouço regulatório, sendo evidente o desafio de redução da inadimplência, especialmente a do setor público.

Apesar dos avanços obtidos no segundo ciclo de revisão tarifária, através da Nota técnica n°. 349/2007-SRE/ANEEL, com uma metodologia para o reconhecimento das receitas incobráveis, os custos com atividades de corte e

reliquação não são cobertos integralmente pela taxa de reliquação cobrada, além de custos com visita de cobrança e assessoria de cobrança, entre outros, não contemplados na empresa de referência (ANEEL, 2007).

Apesar desses pontos controvertidos, os sinais indutores da eficiência produtiva são bastante contundentes. Portanto, constitui um desafio permanente a busca de maximizar o nível de eficiência das concessionárias. Nessa direção, esse trabalho apresenta uma abordagem inovadora de avaliação do risco de inadimplência de consumidores de energia elétrica com o objetivo de maximizar as ações de cobrança e, ao mesmo tempo, minimizar os custos operacionais.

1.2 OBJETIVO DO ESTUDO

Neste estudo apresenta-se uma metodologia, que utiliza técnicas de análise multivariada de dados, para avaliar o comportamento de pagamento e o risco de inadimplência de consumidores de energia elétrica do setor privado.

O modelo é desenvolvido com base na segmentação existente dos consumidores da Coelba, concessionária de distribuição, cuja área de concessão abrange 415 municípios do Estado da Bahia.

Portanto, busca-se desenvolver meios de atuação eficiente na gestão dos valores a receber de uma distribuidora de energia elétrica através da proposição de uma metodologia para avaliação do risco de inadimplência com consumidores, de maneira que, através da análise estatística, seja possível minimizar as perdas decorrentes da inadimplência.

A metodologia proposta busca equacionar, responder ou permitir melhor analisar as seguintes questões sobre o assunto:

Quais os tipos de cobrança mais adequados ao perfil dos consumidores inadimplentes do setor privado?

Qual o risco de inadimplência dos consumidores de energia elétrica pertencentes ao setor privado?

Este estudo busca desenvolver um modelo preditivo de inadimplência para consumidores de energia elétrica através da análise do comportamento de pagamento e cálculo de equações de regressão que explicam a formação da perda esperada ou probabilidade de inadimplência.

A aplicação do modelo busca priorizar, dadas às restrições existentes, as ações de cobrança para os consumidores privados com maiores riscos de tornarem-se inadimplentes, indicando os meios de cobrança com maior efetividade conforme o comportamento de pagamento verificado.

1.3 ORGANIZAÇÃO DO ESTUDO

Buscando atingir os objetivos descritos no item 1.2, este estudo está estruturado em seis capítulos, incluindo a Introdução.

Introdução.

Análise regulatória da inadimplência.

Abordagem teórica sobre risco de inadimplência.

Técnicas de análise multivariada utilizadas.

Estudo de caso.

Conclusão.

No Capítulo I é feita uma análise das principais condicionantes microeconômicas que influenciam a composição da inadimplência no Brasil. É discutido o contexto atual da inadimplência na distribuição de energia elétrica e os custos operacionais associados à gestão da carteira decorrentes do inadimplemento dos consumidores.

O Capítulo II apresenta as características do modelo atual de regulação tarifária associadas com o tratamento regulatório dos custos decorrentes da gestão da inadimplência. São abordados temas como regulação por incentivos, que induz a concessionária a aumentar a eficiência, definição de trajetória de melhoria de

desempenho, tipos de perdas relacionadas à inadimplência e a concentração do risco de inadimplência nas concessionárias de distribuição.

No Capítulo III, são resumidos os principais conceitos de risco de crédito utilizados no mercado financeiro. São discutidos os tipos de riscos e os principais modelos de medição existentes.

No Capítulo IV, analisam-se as técnicas quantitativas utilizadas para alcançar os objetivos da pesquisa. A análise de agrupamentos e regressão logística são as técnicas utilizadas.

O Capítulo V apresenta um estudo de caso para os consumidores privados da Concessionária de Distribuição de Energia Elétrica do Estado da Bahia (Coelba).

Por último, no Capítulo VI é apresentada a conclusão.

Nesse capítulo, buscou-se apresentar uma visão abrangente da inadimplência nas concessionárias do Brasil e do contexto regulatório a que estão submetidas. Foi apresentado também, o objetivo e a organização do estudo.

Portanto, os tópicos apresentados permitem um entendimento sumarizado do conteúdo do estudo.

2 ANÁLISE REGULATÓRIA DA INADIMPLÊNCIA

Seguindo uma tendência mundial de melhoria da eficiência do setor de energia elétrica, o modelo brasileiro de regulação tarifária, adotado desde 1995, fundamenta-se em um modelo de “regulação por incentivos”, que estimula as concessionárias de distribuição a buscar eficiência e redução de custos com a gestão comercial ao longo dos intervalos das revisões tarifárias periódicas (ANEEL, 2003a). Nessa direção, nos próximos itens será apresentado como o órgão regulador determina as metas plurianuais de desempenho dos níveis de perdas devido à inadimplência (ANEEL, 2003b)⁵ e aos riscos de crédito associados à carteira de consumidores, decorrentes desta abordagem.

2.1 O PAPEL DA DISTRIBUIDORA NA ARRECADAÇÃO DE RECURSOS

No atual modelo, a atividade de distribuição passa a ser orientada para o serviço de rede e de venda de energia somente a consumidores com tarifas reguladas pela ANEEL (BRASIL, 2003). Nesse sentido, as concessionárias de distribuição de energia elétrica exercem o importante papel como arrecadoras de recursos para o setor elétrico. Elevados níveis de inadimplência reduzem a capacidade de geração de caixa operacional e influenciam a percepção das agências de classificação de risco de crédito das empresas distribuidoras de energia elétrica (STANDARD & POOR'S, 2004). Assim, em operações para captação de recursos, a exemplo de emissão de debêntures, a classificação de risco (*rating*) atribuída ao tomador é determinante na definição dos juros da operação.

Adicionalmente, a ocorrência da inadimplência provoca perdas com os custos das operações de combate à inadimplência, que não possuem cobertura tarifária, além da perda de receita com faturas de energia consideradas incobráveis.

⁵ Para a Coelba, a trajetória regulatória de redução das perdas devido à inadimplência foi estabelecida na Nota Técnica 052/2003-SRE/ANEEL.

2.2 VISÃO GERAL DA REGULAÇÃO POR INCENTIVOS

Esse modelo de regulação tarifária, adotado para o serviço de distribuição de energia elétrica no Brasil, possui sua origem associada a uma decisão do *Monopolies and Mergers Commission* (MMC) – o órgão responsável pela aplicação da lei antitruste na Inglaterra – em 1982. Em um processo contra determinado ofertante, o MMC recomendou que: i) a variação do preço praticado pela firma fosse limitada à variação de determinado índice de custo de vida, subtraído de um percentual de 1,5%; ii) a implantação da recomendação fosse monitorada; e iii) transcorridos um período de cinco anos, fosse realizada uma revisão dos preços.

Esse modelo foi incorporado ao marco regulatório da Inglaterra em 1984 por *Stephen Littlechild*, após a privatização da *British Telecom* (BT) (ANEEL, 2002).

Nesse contexto, o entendimento do órgão regulador é que essa nova forma de regulação possa evitar que a concessionária possa repassar integralmente os aumentos de custos para a tarifa, estimular a melhoria da produtividade e da eficiência, os investimentos e a inovação tecnológica. Como contrapartida, a regulação por incentivos, busca assegurar a apropriação de ganhos de produtividade do concessionário durante o intervalo das revisões tarifárias periódicas (ANEEL, 2003a; VILLELA; MACIEL, 1999). Portanto, com o objetivo de aumentar a eficiência produtiva das concessionárias de distribuição de energia elétrica, desde 1995, o modelo de regulação econômica brasileiro tem sido aperfeiçoado para incentivar os agentes regulados a maximizar a eficiência produtiva.

Na regulação “incentivada”, as soluções normalmente utilizadas são: regulação por preço (*price cap*), regulação por padrão de comparação (*yardstick competition*) e regulação da qualidade (ARAÚJO, 1997).

De acordo com Possas, Pondé e Fagundes (2004), o regime tarifário *price cap* compreende uma regra de reajuste por índice público de preços, acompanhada de previsão de redução de custos por aumento de produtividade, com o objetivo de estimular, de forma muito simples e transparente, a busca de aumento de eficiência microeconômica. A fórmula típica:

$$p = IPC - X$$

onde: IPC é um índice de preços (originalmente ao consumidor, mas isso não é estritamente necessário);

e X: um redutor de produtividade predeterminado.

O *price cap*, segundo Pires e Piccinini, (1998, p.18), é um mecanismo tarifário que se constitui na definição de um preço-limite para os preços médios da empresa, corrigido de acordo com a evolução de um índice de preços ao consumidor, menos um percentual equivalente a um fator X de produtividade, para um período pré-fixado, normalmente, entre três e oito em anos,

Outro esquema de regulação que busca incentivar a redução de custos e preços é a “regulação por padrão de comparação” (*yardstick regulation*), também chamada “competição por padrões” (*yardstick competition*). Nesse esquema, o órgão regulador tem acesso aos balanços anuais das diversas concessionárias, e fixa preços para o ano seguinte baseado nas empresas mais eficientes, com certa tolerância. Nesse esquema, supõe-se que não haja colusão entre as firmas e que suas estruturas de custos sejam comparáveis, de modo que os custos das firmas sejam agregados em “custos-padrão” (ARAÚJO, 1997).

O regime tarifário baseado no desempenho (*yardstick competition*) é fundamentado na introdução de incentivos à maior eficiência pela redução de assimetrias de informações quando há várias empresas reguladas, como por exemplo, no setor de distribuição de energia elétrica. O desempenho da firma regulada é aferido pela comparação com uma referência pré-estabelecida, um *benchmark*, que induza ao aumento de produtividade e redução de custos (POSSAS; PONDÉ; FAGUNDES, 2004).

Araújo (1997) conceitua “regulação da qualidade” como esquemas que buscam garantir a qualidade dos bens ou serviços (expressa em indicadores observáveis pelo regulador) através de exigências ou de incentivos. Isto pode ser encontrado em associação com regulação do preço. A idéia desta família de esquemas é compensar impactos negativos da regulação do preço sobre a

qualidade, sem exigir o mesmo volume de informação da regulação por custo de serviço. Araújo (1997) também afirma que, apesar da idéia ser interessante, há poucos estudos a respeito, contudo, recomenda ver Laffont e Tirole (1993), capítulo

4.

Segundo Araújo (1997), a regulação da qualidade normalmente está associada à regulação do preço para contrabalançar com o viés potencial desta última contra a qualidade. Por exemplo, a busca por menores custos de operação e manutenção poderia comprometer os índices de duração e frequência das interrupções no fornecimento de energia elétrica. A idéia geral da regulação da qualidade é compensar, através de multas, possíveis impactos negativos da regulação do preço sobre a qualidade, sem exigir o mesmo volume de informação como em outros regimes tarifários.

No Brasil, as atuais regras jurídicas e econômicas pertinentes ao regime tarifário dos contratos de concessão do serviço público de distribuição de energia elétrica constituem uma vertente do regime de regulação por incentivos. Assim, o art. 14 da Lei nº 9.427, de 26 de dezembro de 1996, estabelece que o regime econômico e financeiro da concessão de serviço público de energia elétrica compreende “IV - apropriação de ganhos de eficiência empresarial e da competitividade” (ANEEL, 2003b).

Portanto, é um pressuposto desse modelo o estímulo ao concessionário para maximizar os lucros através da busca por custos cada vez menores, uma vez que lhe é permitido apropriar-se dos ganhos obtidos durante os intervalos entre as revisões tarifárias periódicas. Na regulação por incentivos, são estabelecidas metas que induzem as empresas reguladas a atingir desempenhos desejáveis, através da concessão temporária de algum direito. Os preços são reajustados para baixo quando a taxa de retorno excede uma taxa limite, mas o ajuste é parcial para que a empresa fique com parte do lucro realizado (PINTO JR.; SILVEIRA, 2004, p. 11). Por outro lado, o concessionário também absorverá o risco de perdas por operar com custos ineficientes, ou seja, custos acima dos custos da “empresa de referência”, que é uma firma hipotética criada pelo Agente Regulador para comparação com a firma real e com objetivo de estimular a competição em custos.

Deste modo, o modelo de regulação tarifária vigente assegura à empresa concessionária o direito de apropriar-se dos ganhos de produtividade durante o intervalo regulatório⁶. Ao final desse intervalo regulatório, que normalmente tem duração entre três e oito anos, o órgão regulador fará a revisão tarifária periódica e transferirá os ganhos de produtividade para o consumidor. Na hipótese inversa, ou seja, havendo perda de produtividade, os custos adicionais serão absorvidos pelo concessionário e não serão repassados para a tarifa.

2.3 REGULAÇÃO DA INADIMPLÊNCIA

O atual regime tarifário do serviço de distribuição de energia elétrica brasileiro é baseado em preços-teto, cujo objetivo principal é a eficiência na prestação do serviço e a modicidade tarifária (ANEEL, 2003b). No entanto, o aparato de incentivos presentes na regulação tarifária atual assume a feição de um modelo de regulação por incentivos, que é uma variante da regulação por preços-teto (VILLELA; MACIEL, 2004, p.13). Em resumo, no modelo brasileiro de regulação por incentivos estão presentes esquemas de regulação tarifária de regulação do preço (*price cap*), regulação por padrão de comparação (*yardstick competition*)⁷ e de regulação da qualidade (ARAÚJO, 1997, p. 15). Assim, após uma primeira revisão tarifária periódica, apoiada em auditoria de informações fornecidas pelo concessionário, o regulador determina os “custos operacionais eficientes” (ANEEL, 2003b).

A avaliação dos custos operacionais eficientes pode ser prejudicada pela assimetria de informação, uma vez que o agente regulado é quem detém a melhor informação sobre a sua concessão. Dessa forma, para aumentar o nível de eficiência econômica, o regulador define uma “trajetória regulatória” de redução

⁶ Período entre as revisões tarifárias periódicas que dura entre três e oito anos. No caso da Coelba, o intervalo regulatório é de cinco anos.

⁷ A *yardstick competition* é uma forma de regulação através de incentivos, também conhecida como regulação de desempenho, adotada nos casos de monopólio natural. Este instrumento procura introduzir estímulo à redução de custos entre as empresas, reduzir as assimetrias de informação existentes e estimular maior eficiência econômica (ARMSTRONG; COWAN; VICKERS, 1994 apud PIRES; PICCININI, 1998, p.28). De fato, o regulador cria uma empresa de referência hipotética para comparação com a firma real e fomentar a competição em custos.

da inadimplência, na qual custos superiores aos estabelecidos não serão repassados para a tarifa.

Neste ponto, é importante ressaltar a forte relação entre as perdas com devedores inadimplentes e as perdas comerciais que decorrem da ação fraudulenta do consumidor. Existe um risco potencial de deslocamento das perdas entre estes dois tipos, ou seja, ações de combate às perdas comerciais podem transformar fraudadores em inadimplentes. Por exemplo, uma fatura de fraude não recebida passa a compor a provisão com créditos de liquidação duvidosa e posteriormente ser reconhecida como perda financeira. Assim, para decidir a melhor estratégia de atuação, a empresa pode analisar tanto a trajetória definida pelo regulador para as perdas totais, quanto para a inadimplência.

Pode-se inferir que o regulador, ao fixar uma trajetória regulatória exclusiva para as perdas por inadimplência, considerou que a concessionária possui capacidade de gestão sobre a totalidade da inadimplência dos consumidores (ANEEL, 2003b, p. 16). Portanto, o controle da inadimplência e das perdas comerciais constitui duas frentes de trabalho que devem ser controladas simultaneamente, para alcançar os patamares mínimos de eficiência produtiva. Assim, diante de uma “trajetória regulatória” estabelecida, uma concessionária que consegue atingir níveis de desempenho superiores pode, portanto, apropriar-se dos ganhos de produtividade.

2.4 EFEITOS DA INADIMPLÊNCIA SOBRE A PRODUTIVIDADE

Neste trabalho, a medição da inadimplência de consumidores orienta-se através da base legal para delimitação dos valores das receitas irrecuperáveis estabelecida no artigo 340, do Regulamento do Imposto de Renda (RIR) e nas instruções contábeis do *Manual de Contabilidade do Serviço Público de Energia Elétrica* (ANEEL, 2002), item 6.3.2. Os valores estimados das perdas são contabilizados em uma provisão para créditos de liquidação duvidosa - PCLD, que representa uma despesa que exclui do ativo os valores considerados de difícil recebimento.

A inadimplência significa que parte da energia faturada aos consumidores não foi recebida. Os efeitos da inadimplência compreendem também os recursos utilizados para as operações de cobrança. Com esse custo adicional, a concessionária necessita de um volume maior de receita para cobrir os custos de gestão das perdas devido à inadimplência.

Portanto, os custos operacionais envolvidos na gestão da inadimplência não são reconhecidos adequadamente na revisão tarifária periódica, tampouco os custos de capital devido ao aumento da necessidade de capital de giro e as perdas com devedores incobráveis reconhecidas no balanço patrimonial da empresa concessionária. Esses fatores reduzem os valores repassados para as tarifas admitidos como custos operacionais eficientes

2.4.1 A base de cálculo da inadimplência regulatória

A metodologia utilizada na revisão tarifária considera o faturamento do ano teste futuro para projetar todos os custos que devem ser remunerados pela receita a ser auferida no mercado da concessionária. Em oposição, a base de cálculo utilizada para definir a inadimplência regulatória é o faturamento verificado no ano anterior à revisão tarifária periódica, de maneira que se verifica uma sub-avaliação dos valores utilizados como base para o cálculo das despesas com inadimplência. No caso da Coelba, a receita de fornecimento verificada em 2002 (estimada para o ano-teste) foi de R\$ 1.430.836.860,00, enquanto que a receita total requerida para 2003 (estimada para o ano-teste futuro) fora de R\$ 1.920.209.159,42 (Nota Técnica n°. 052/2003-SRE/ANEEL (ANEEL, 2003a))⁸. Assim, verifica-se que 489 milhões de reais ficaram de fora da referida base de cálculo, sobre a qual estaria o montante de inadimplência.

Adicionalmente, ao adotar uma trajetória regulatória sob a forma de um percentual do faturamento anual, o Agente Regulador exclui da base de cálculo os valores do Imposto sobre Circulação de Mercadorias e Serviços (ICMS) faturado e a receita projetada para o ano teste futuro (ANEEL, 2003b, p. 20).

⁸ Valores revisados na Nota Técnica N°. 154/2003-SRE/ANEEL.

O ICMS decorre de uma obrigação tributária prevista em lei. Assim, esse tributo, uma vez faturado, é recolhido aos cofres públicos sem que sejam efetivamente reconhecidos como perdas devido à inadimplência. De forma similar, verifica-se que a determinação legal para o cálculo da provisão de devedores duvidosos, prevista no regulamento do imposto de renda, não discrimina os valores das perdas em mercadorias e impostos, como faz o Agente Regulador na apuração dos custos com inadimplência (BRASIL, 1999)⁹.

A provisão de perdas corresponde ao valor total da fatura, inclusive o ICMS, logo, por analogia, é legalmente aceitável considerá-lo como parte integrante do custo da energia elétrica fornecida. Assim, a trajetória regulatória de redução da inadimplência deveria considerar a receita bruta de fornecimento, inclusive o ICMS.

Para o segundo ciclo de revisão tarifária, houve alterações na metodologia que define a inadimplência regulatória, agora denominada “receitas irrecuperáveis”.

Para o primeiro ciclo de revisão tarifária, foram adotados, como receitas irrecuperáveis para as distribuidoras de energia elétrica, percentuais de 0,5%, 0,4%, 0,3% e 0,2% do faturamento bruto (sem ICMS), respectivamente, para o primeiro, segundo, terceiro e quarto (e quinto, se houver) ano do ciclo tarifário, para todas as concessionárias.

Para o segundo ciclo de revisão tarifária periódica, a Resolução Normativa n.º 234/2006 definiu que o percentual máximo regulatório para as receitas irrecuperáveis a serem reconhecidas para as distribuidoras de energia elétrica será calculado observando os seguintes procedimentos:

- a) Levantamento e análise dos totais de receitas irrecuperáveis efetivamente incorridas pelas empresas;
- b) Agrupamento das empresas por conjuntos de similaridade;

⁹ Regulamento do Imposto de Renda (RIR), Decreto Lei n.º. 3.000, de 26/03/1999, artigo 340.

- c) Definição dos percentuais individuais das empresas com estabelecimento de metas para o próximo.

Através da Nota Técnica n.º. 349/2007-SRE/ANEEL foi proposto um modelo para regulação de receitas irrecuperáveis. De uma maneira geral, a metodologia consiste na obtenção de um índice de inadimplência definido a partir da média da inadimplência verificada das empresas com características assemelhadas, inclusive da própria. Outra regra possível é a da mediana, muito recomendada quando há forte assimetria na distribuição dos dados.

Conforme descrito na mesma Nota Técnica n.º. 349/2007-SRE/ANEEL,

A metodologia a ser empregada consiste em classificar as empresas por grupos de similaridade (sócio-econômica) e definir o valor de referência de receitas irrecuperáveis para o grupo de acordo com uma regra associada aos valores do grupo, de forma a garantir o incentivo adequado para a busca de um nível ótimo de inadimplência. (ANNEEL, 2007, p.1).

Na mesma Nota Técnica, também está descrito que a definição dos grupos de similaridade será feita através do índice construído na Nota Técnica n.º. 348/2007-SRE/ANEEL, relativa à metodologia de perdas não-técnicas. Na medida em que o índice permite diferenciar de forma adequada as áreas de concessão quanto aos seus indicadores econômicos, e devido à relação estreita entre os temas, o procedimento é coerente.

Os resultados apresentados nesse documento, considerados adequados pela ANEEL, apresentam a classificação das empresas em 3 grupos que levaram em conta o ordenamento das empresas no ranking obtido da Nota Técnica n.º. 348/2007-SRE/ANEEL e as diferenças regionais entre áreas de concessão. As empresas menores foram consideradas em um grupo a parte devido à maior facilidade destas na gestão dos seus níveis de inadimplência.

Assim, o nível de inadimplência de referência das distribuidoras foi de 0,90%, 0,60% e 0,20%, respectivamente para as distribuidoras do Grupo 1, Grupo 2 e Grupo 3. Através desses percentuais obtêm-se os valores de receitas irrecuperáveis de referência para cada grupo de empresas.

2.4.2 Perdas por inadimplência de consumidores

Para a inadimplência dos consumidores, expressa pela “provisão de créditos de liquidação duvidosa (PCLD)”, o Agente Regulador adotou um conceito de custo operacional “transitivo” da Empresa de referência. Trata-se da definição de uma trajetória de melhoria contínua do desempenho, sob a forma de um percentual do importe faturado verificado no ano da revisão tarifária periódica. Essa ação do regulador visa evitar que os consumidores em situação regular paguem pelos consumidores inadimplentes (ANEEL, 2003b, p. 18).

É válido o entendimento de que os consumidores adimplentes não devem pagar para manter elevados índices de inadimplência. No entanto, as perdas devido à inadimplência são inerentes à atividade empresarial. De fato, não existe “perda zero”, como definida na trajetória regulatória prevista na Nota Técnica 052/2003-SRE/ANEEL, tampouco uma solução simples que reduza significativamente os patamares de inadimplência.

Portanto, a definição de uma metodologia de tratamento das receitas irrecuperáveis, através da Nota Técnica n.º. 349/2007-SRE/ANEEL, que considera os aspectos de concessão, representa avanço no reconhecimento de custos com inadimplência, sobretudo concessões localizadas em regiões socialmente desfavorecidas, onde a dívida é fortemente distribuída em faturas de baixo valor.

2.4.3 Custos de captação do capital de giro

No Brasil, em uma distribuidora de energia elétrica, os créditos com consumidores não possuem garantias reais. Portanto, acumulam riscos de crédito associados à inadimplência, à degradação do crédito e à concentração do crédito em grupos reduzidos de consumidores (Res. n.º. 456)¹⁰ (ANEEL, 2000 apud SAUNDERS, 2000). O perfil da carteira de valores a receber de consumidores e o desempenho da gestão da cobrança são elementos formadores do risco de crédito

¹⁰ Resolução normativa n.º.456, de 29/11/2000: Estabelece as disposições atualizadas e consolidadas, relativas às condições gerais de fornecimento de energia elétrica, a serem observadas na prestação e utilização do serviço público de energia elétrica, tanto pelas concessionárias e permissionárias, quanto pelos consumidores.

medidos por agências classificadoras de risco, podendo elevar o custo de capital para a empresa tomadora em função da diminuição da capacidade de geração de caixa.

O risco de crédito da empresa é um importante fator para determinação da capacidade de endividamento e cobertura de empréstimos. Desse modo, a manutenção de elevados valores a receber de consumidores implica na redução do fluxo de caixa líquido e na provável captação de recursos para cobrir a necessidade de capital de giro ou para fazer investimentos (STANDARD & POOR´S, 2004, p.4). As despesas financeiras decorrentes dessas operações não são consideradas para efeito de reajuste tarifário.

Assim, se pode afirmar que o mau desempenho da cobrança tanto eleva os custos operacionais e as perdas com créditos de liquidação duvidosa (PCLD), quanto os custos com captação de recursos para manutenção do capital circulante e para investimentos.

2.4.4 Custos operacionais da cobrança

Após a apresentação das faturas aos consumidores, inicia-se o que pode ser denominado de “corrida contra o tempo”. A distribuidora passa a incorrer em diversos custos operacionais durante a cobrança dos consumidores em atraso. Na Nota Técnica nº. 052/2003-SRE/ANEEL, não são considerados os custos com cartas de cobrança previstas na Resolução 456/2000/ANEEL, tampouco os custos com cortes e religamentos. Mais uma vez, o entendimento do regulador é que esses serviços devem ser pagos pelos clientes que geram os respectivos custos à concessionária, mediante a cobrança de taxas específicas. Para que o entendimento do regulador seja operacionalizado sem prejuízos para a distribuidora, faz-se necessário o estabelecimento de taxas que cubram efetivamente o custo de cada atividade.

No caso da Coelba, mesmo após a emissão da carta de cobrança, persiste um elevado volume de valores não recebidos, agravado pela forte dispersão dos créditos em consumidores com baixo consumo mensal. Frente a esses fatores, via

de regra, o volume de atividades de cobrança é superior à capacidade operacional da empresa, acentuando o risco de crédito.

Neste capítulo, apresentou-se o segmento de distribuição de energia elétrica sob o aspecto da regulação por incentivos e inadimplência. A distribuição de energia elétrica, considerada “monopólio natural”, está submetida a um modelo de regulação econômica baseado em preços-teto e incentivos à eficiência na direção de custos operacionais eficientes. Verificou-se que a assimetria de informação é motivo mais que suficiente para o Agente Regulador estabelecer uma “trajetória regulatória” de redução da inadimplência dos consumidores. O ponto de atenção é a definição de metas de desempenho aparentemente sem uma base científica ou conhecimento empírico para o primeiro ciclo de revisão tarifária. Para segundo ciclo houve importantes avanços ao considerar as características das distribuidoras e aspectos sócio-econômicos da concessão.

Adicionalmente, não se considerou, de forma apropriada, todos os custos incorridos na gestão comercial, a exemplo de custos de cobrança a consumidores e a totalidade das despesas com dívidas incobráveis. O entendimento do Agente Regulador é que os serviços de corte e religação de unidades consumidoras “devem ser pagos pelos consumidores que geram os respectivos custos à concessionária, mediante a cobrança de taxas específicas e, portanto, não devem ser incluídos na tarifa que pagam os clientes em condição regular” (ANEEL, 2003b, p. 20). Desse modo, os custos com corte e religação, mesmo não sendo incorporados à tarifa, deveriam cobrir os custos reais incorridos pelas distribuidoras.

Assim, principalmente em função da assimetria de informação, a definição de “custos operacionais eficientes” ainda representa um desafio para determinação da tarifa adequada. A análise do modelo vigente de regulação por incentivos, permitiu inferir que o conjunto de medidas de incentivo à eficiência econômica, a exemplo da “empresa de referência”, não considerou adequadamente as peculiaridades do mercado de cada área de concessão e podem impedir a empresa de obter o desempenho fixado pelo regulador (perdas comerciais, inadimplência).

Portanto, diante da trajetória regulatória da inadimplência, a distribuidora deve estabelecer políticas adequadas para minimizar os riscos e maximizar o desempenho retendo os ganhos de eficiência alcançados durante o intervalo regulatório.

3 ABORDAGEM TEÓRICA SOBRE RISCO DE INADIMPLÊNCIA

No atual modelo, o segmento de distribuição é o principal arrecadador de recursos, além de absorver a maior parte do risco de crédito do setor elétrico.

Frente a um tratamento regulatório que induz a busca de custos operacionais eficientes e que determina limites de custos que serão repassados ao consumidor, o risco de perdas de créditos é acentuado. Assim, urge a necessidade de aprofundar o conhecimento da carteira de créditos a receber de consumidores e os riscos associados. Como objetivo específico, cabe à distribuidora estabelecer políticas e estratégias adequadas para eliminar ou minimizar os riscos de crédito da carteira de consumidores.

Os principais tipos de riscos de crédito associados à carteira de valores a receber de consumidores em uma distribuidora de energia elétrica podem ser definidos como riscos associados à idade da inadimplência, à degradação do crédito pela perda de valor, aos tipos de faturas e à concentração do crédito em grupos reduzidos de consumidores (CAOQUETTE; ALTMAN; NARAYANAN, 1999).

É importante enfatizar o fato de que os créditos do fornecimento de energia elétrica não possuem garantias reais. Portanto, os riscos associados a esses créditos devem ser objetos de análise e identificação dos principais atributos que interferem na sua correta mensuração.

Literalmente, uma concessionária de distribuição de energia elétrica vende a prazo. Diferentemente das empresas submetidas ao regime de mercado, que possuem a prerrogativa de analisar e decidir incondicionalmente sobre a concessão de crédito, as empresas sujeitas ao regime de serviço público obrigam-se a conceder o crédito ao consumidor, independentemente do risco de crédito associado.

De acordo com o artigo 4º da Resolução nº. 456/2000/ANEEL, para uma distribuidora de energia elétrica, resta apenas a possibilidade de condicionar a ligação, religação, alterações contratuais, aumento de carga ou contratação de fornecimentos especiais, solicitados por quem tenha quaisquer débitos no mesmo ou em outro local de sua área de concessão, à quitação dos referidos débitos (Res.

nº 456) (ANEEL, 2000). Para tanto, o cadastro dos consumidores deve estar completo e atualizado. A qualidade do cadastro é importante fator auxiliar nas ações de cobrança.

Desse modo, a avaliação do risco de inadimplência, enquanto uma medida de incerteza de pagamento assume cada vez mais importância na agenda de uma distribuidora de energia elétrica. Esse cenário obriga a empresa a reduzir a inadimplência dada às várias restrições de ordem regulatória, operacional e financeira.

O risco de crédito decorre de uma transação financeira contratada entre um fornecedor de fundos e um usuário desses fundos. Antes do surgimento das empresas operando em um regime capitalista, as famílias e o governo suportavam a maior parte do risco de crédito. A formação das sociedades por ações possibilitou a criação de empresas capazes de suportar o risco econômico (CAOUILLE; ALTMAN; NARAYANAN, 1999).

O crédito é a expectativa de recebimento de um determinado valor em dinheiro em uma data previamente definida. A simples concessão de crédito implica em risco de crédito. Assim, Caouette, Altman e Naraynan (1999, p.1) ensinam que o risco de crédito é a possibilidade de que a expectativa de recebimento não se realize. Para Crouhy, Galai e Mark (2004, p.34), o risco de crédito é o risco de que uma mudança na qualidade do crédito afetará o valor da carteira.

Quineche Reyna (2004) apresenta o conceito de risco de crédito associado à mensuração de fatores que determinam o grau de exposição ao risco.

Risco de crédito é a medida numérica da incerteza com relação ao recebimento de um valor contratado (ou compromissado), a ser pago por um tomador de um empréstimo, contraparte de um contrato ou emissor de um título, descontadas as expectativas de recuperação e realização de garantias.

Os fatores determinantes do risco de créditos são i) Risco de Inadimplência, ii) Risco de Degradação de Crédito, iii) Risco de Degradação de Garantias, iv) Risco Soberano, v) Risco de Financiador, vi) Risco de Concentração (crédito). De

forma objetiva, Brasileiro (2002, p.2), apresenta os conceitos dos fatores do risco de crédito.

a) Risco de Inadimplência.

Este pode ser definido como o risco de perda pela incapacidade de pagamento do tomador de um empréstimo. Exemplos: i) não pagamento de faturas de energia elétrica, de juros e/ou principal de crédito pessoal, empréstimos para pessoa jurídica, cartão de crédito, *leasing*, etc.; ii) não pagamento de juros e/ou principal de títulos de renda fixa (nacionais ou internacionais, públicos ou privados) pelo emissor.

b) Risco de Degradação de Crédito

Pode ser definido como o risco de perdas pela degradação da qualidade creditícia do tomador de um empréstimo, levando a uma diminuição no valor de suas obrigações. Exemplos: i) perdas em títulos soberanos e/ou corporativos pela redução do *rating* do país emissor.

c) Risco de Degradação das Garantias

Define-se como o risco de perdas pela degradação da qualidade das garantias oferecidas por um tomador de um empréstimo. Exemplos: i) empréstimos cujas garantias não mais existam; ii) depreciação no valor das garantias depositadas em bolsas de derivativos.

d) Risco Soberano

Este é definido como o risco de perdas pela incapacidade de um tomador de um empréstimo em honrar seus compromissos em função de restrições impostas por seu país sede. Exemplos: i) transações que envolvam transferências internacionais de títulos ou de câmbio.

e) Risco de Financiador

Definido como o risco de perdas por inadimplência do financiador de uma transação, potencializada quando o contrato não contempla acordo de liquidação

por compensação de direitos e obrigações (*netting agreement*). Exemplos: i) acordo de recompra (*repurchase transactions*) que não contemplem a liquidação (*netting*) de direitos ou obrigações.

f) Risco de Concentração (crédito)

Este é definido como o risco de perdas em decorrência da não diversificação de risco de crédito de investimentos. Exemplos: i) concentrar empréstimos em poucos setores da economia, classes de ativos, etc.; ii) possuir parte substancial dos passivos de um devedor (por exemplo, um emissor de debêntures).

Uma operação de crédito está associada a fatores de incerteza quanto à realização do pagamento pelo tomador. No setor de serviços públicos, especificamente em empresas de energia elétrica, gás canalizado, telefonia e água e saneamento, o serviço ou produto é entregue mediante um compromisso de pagamento em data futura. Na qualidade de concessionária ou permissionária do serviço público, esse tipo de empresa obriga-se operar em obediência aos marcos regulatórios definidos pelo Agente Regulador. Desse modo, aceitam o risco de crédito com seus consumidores ou assinantes sem garantias prévias. O tópico a seguir, trata do elemento-chave do risco de crédito: o risco de inadimplência.

3.1 RISCO DE INADIMPLÊNCIA COM CONSUMIDORES DE ENERGIA ELÉTRICA

A inadimplência ocorre quando o tomador não quer ou não pode cumprir com suas obrigações contratuais (CROUHI; GALAI; MARK, 2004). A inadimplência é o elemento-chave na avaliação do risco de crédito, sejam tomados individualmente ou em carteiras (CAOUILLE; ALTMAN; NARAYANAN, 1999, p. 221). Do mesmo modo que uma operadora de cartão de crédito concede o crédito para pagamento futuro, a distribuidora (DEE), ao antecipar o fornecimento de energia elétrica ao consumidor final, assume o risco de inadimplência. Essa imposição regulatória pode ser conceituada como uma concessão de um crédito rotativo não-garantido pela DEE ao consumidor de energia elétrica. No caso de inadimplemento do consumidor, a cobrança do valor devido ocorre em média, 77 dias após o vencimento da fatura (SAMP - SEMINÁRIO ABRADEE..., 2005).

Para o serviço público de distribuição de energia elétrica, esse mesmo Agente Regulador, estabelece regras indutoras de eficiência produtiva ao limitar os níveis de inadimplência remunerados através da tarifa de energia elétrica (Nota Téc. 030) (ANEEL, 2003b). Esse dispositivo cria condições para que a DEE busque minimizar a inadimplência e os custos operacionais relacionados. Nesse arcabouço, as despesas com inadimplência incorridas pela DEE acima do limite estabelecido pelo Agente Regulador são absorvidas sem qualquer contrapartida tarifária.

Na pesquisa realizada em 2006 pela Associação Brasileira de Distribuidoras de Energia Elétrica (ABRADEE) apresenta-se um panorama sobre a inadimplência no setor de DEE brasileiro. Em que pese o esforço despendido para o controle da inadimplência, níveis elevados de inadimplência persistem. São sobremaneira elevados em comparação com a inadimplência regulatória definida na empresa de referência. Na Tabela 2, a seguir, é apresentado um resumo da inadimplência nas principais concessionárias de distribuição de energia elétrica do Brasil.

Tabela 2 - Inadimplência (%) - Contas Vencidas/Faturamento Acumulado nos últimos 12 meses. (dados com PDD). Amostra de 15 distribuidoras = 62% do mercado

Empresa	30/jun/03		30/jun/04		30/jun/05		30/jun/06	
	Até 90 dias	Total						
AES SUL	2,15	6,68	2,30	7,23	2,05	7,64	1,92	5,32
AMPLA	5,29	22,52	4,57	19,70	4,30	18,42	4,02	15,77
BANDEIRANTE	2,47	4,49	3,21	5,60	3,09	4,08	3,04	5,02
CELESC	3,27	7,12	2,94	6,36	2,77	8,32	2,81	9,22
CELPA	4,30	7,83	4,31	7,31	3,52	5,67	3,94	7,95
CELPE	8,36	12,67	6,39	9,96	6,91	11,52	6,20	12,37
COELBA	6,17	9,73	4,36	6,29	5,31	6,98	3,68	5,19
COPEL	3,10	4,89	2,42	3,59	2,27	4,48	2,22	3,60
COSERN	5,90	13,90	4,14	9,65	5,48	9,04	5,78	8,79
CPFL	3,15	3,95	2,43	3,32	2,74	3,94	2,73	3,71
ELEKTRO	2,20	3,50	2,64	3,55	2,67	3,89	2,56	3,97
ELETROPAULO	2,32	6,05	2,34	6,50	2,64	6,40	2,26	4,55
LIGHT	7,13	20,61	6,61	20,17	6,55	18,33	5,21	13,27
PIRATININGA	2,12	3,34	2,09	2,76	2,30	3,53	2,41	3,60
RGE	3,24	7,61	2,87	6,84	2,54	6,31	1,97	5,74
MOSTRA	3,72	8,45	3,34	7,65	3,45	7,69	3,16	6,79

Fonte: CVM, ABRADEE (2006).

3.2 MODELOS DE MEDIÇÃO DO RISCO DE INADIMPLÊNCIA

A literatura envolvendo análise quantitativa de carteiras de consumidores, com o objetivo de minimizar o risco de inadimplência, tem sido focada em estudos de previsão de insolvência. O prognóstico correto da insolvência futura pode

auxiliar as instituições de crédito a evitar perdas e reduzir o custo do empréstimo para os bons tomadores. De acordo com Caouette, Altman e Naraynan (1999) e Crouhy, Galai e Mark (2004), é grande a amplitude de aplicação dos sistemas de administração de risco de crédito ao consumidor, a exemplo de crédito parcelado ao varejo, financiamento e *leasing* para automóveis e outros bens duráveis, além de empréstimos em geral. Ainda de acordo com os autores, esses sistemas empregam uma variedade de técnicas: julgamento humano, *credit scoring*, *behavior scoring*, análises de árvores de decisões, programação matemática e mais recentemente redes neurais. Saunders (2000) aponta três modelos de medição do risco de inadimplência, que compreendem abordagens tradicionais e novas: sistemas especialistas, sistemas de classificação e sistemas de pontuação de crédito. Uma discussão mais detalhada desses modelos é apresentada em Caouette, Altman e Naraynan (1999).

Em um sistema especialista, exige-se um conhecimento especializado por parte do profissional responsável pela análise do crédito, uma vez que seu julgamento subjetivo e atribuição de pesos aos fatores analisados são determinantes na decisão de conceder ou não ou crédito. Um dos sistemas especialistas mais comuns, denominado "os cinco C". Nessa abordagem, o especialista analisa e atribui pesos subjetivamente a cinco fatores-chave e obtém dados para uma decisão de crédito. Esses cinco fatores são: i) Caráter: mede a reputação do tomador principalmente através do tempo como cliente e histórico de pagamentos; ii) Capital: mede a participação do capital próprio e o endividamento do tomador; iii) Capacidade: mede a capacidade de pagamento e a volatilidade dos ganhos do tomador; iv) *Collateral* (garantias reais): mede o risco de exposição do empréstimo considerando os bens e direitos dado pelo tomador como garantia real em caso de inadimplência; v) Ciclo ou condições econômicas: que mede o estado do ciclo do negócio ao qual o tomador está inserido. Fatores como sazonalidade, tecnologia e condições competitivas representam elementos importantes na determinação do risco de inadimplência. Além dos "cinco C" é comum o especialista levar em conta o nível atual e a tendência das taxas de juros, haja vista que a relação entre o nível das taxas de juros e o retorno esperado sobre um investimento é não-linear (STILIGTZ; WEISS, 1981 apud SAUNDERS, 2000, p.8).

Em um sistema especializado as decisões de crédito refletem o julgamento pessoal de um profissional treinado a respeito da capacidade de pagamento que o tomador possui. Por esse motivo, esse tipo de sistema sofre de problemas de consistência, quando é necessário distinguir os diferentes tipos de tomadores. Além disso, é afetado pela própria subjetividade na atribuição dos pesos aos fatores-chave.

Os sistemas de classificação para empréstimos provêm informações codificadas sobre a capacidade creditícia de emitentes de títulos corporativos ou governamentais e tomadores de empréstimo. De acordo com Saunders (2000), bancos e reguladores utilizam esse tipo de sistema com o objetivo de avaliar a adequação de suas reservas para perdas decorrentes de empréstimos. Essa tarefa é desenvolvida por agências de *rating* com a função de informar os investidores sobre a probabilidade de não pagamento e qual o nível de cobertura do risco de crédito a ser provisionado.

Com a ascensão dos mercados de capitais como fonte primária de crédito, os sistemas de classificação de crédito, através das agências de *rating* passaram a desempenhar um importante papel na administração do risco de crédito. De acordo com Caouette, Altman e Naraynan (1999), por esse motivo, sua utilização não ficou restrita aos investidores e foi incorporada em regulamentações governamentais como meio de assegurar que bancos, fundos de pensão e companhias de seguros mantenham as carteiras de renda com suficiente qualidade creditícia.

De uma maneira geral, Saunders (2000) resume o mecanismo de um sistema de classificação de crédito: enquadrar uma carteira de empréstimo existente em categorias, de acordo com o modelo adotado pela agência de *rating* e indicar a reserva para perda decorrente do empréstimo. Um exemplo de um sistema de classificação de empréstimos adotado pelas duas principais agências de *rating*, a *Moody's Investors Service* e a *Standard & Poor's (S&P)*, é mostrado na Tabela 3.

Tabela 3 - Um exemplo de sistema de classificação de empréstimos: símbolos de *rating* de créditos seniores de longo prazo

Ratings de grau de investimento			Ratings de grau especulativo		
S&P e outras	Moody's	Interpretação	S&P e outras	Moody's	Interpretação
AAA	Aaa	Mais Alta qualidade; extremamente forte.	BB+	Ba1	Provável cumprimento de obrigações; incerteza corrente.
AA+	Aa1	Alta qualidade	BB	Ba2	Obrigações de alto risco
AA	Aa2		BB-	Ba3	
AA-	Aa3		B+	B1	
A+	A1	Forte capacidade de pagamento	CCC+	B2	Vulnerabilidade presente à inadimplência.
A	A2		CCC	B3	
A-	A3		CC-	Caa1	
BBB+	Baa1	Capacidade de pagamento adequada	CC	Caa2	Em falência ou inadimplência ou com outros problemas.
BBB	Baa2		C	Caa3	
BBB-	Baa3		D	Ca	

Fonte: Cantor e Parker (1994), *apud* Caouette, Altman e Naraynan (1999)

O procedimento utilizado pela Standard & Poors é um bom exemplo de como as agências de *rating* avaliam um título industrial e atribuem índices de classificação. Essa agência se concentra em duas áreas básicas: i) Risco do negócio: avalia as características do setor, a qualidade da administração do negócio e a posição competitiva da indústria em áreas como marketing, tecnologia e produtividade; ii) Risco financeiro: avalia as características financeiras, a política financeira, a lucratividade, a estrutura de capital, o fluxo de caixa e a flexibilidade financeira. Cada área avaliada é ponderada através da atribuição de pesos.

O terceiro modelo de medição do risco de inadimplência apontado por Saunders (2000), refere-se aos sistemas de *score* de crédito (*credit scoring*). Um *credit score* é número atribuído à qualidade de crédito de um tomador para auxiliar a decisão de conceder ou não um empréstimo ou financiamento. O sistema de *credit scoring* de crédito é largamente utilizado, abrangendo desde crédito ao consumidor até empréstimos comerciais. O princípio desse tipo de sistema é o mesmo utilizado em outros sistemas, ou seja, buscam identificar previamente as variáveis preditoras da inadimplência e inferir uma probabilidade de que o tomador passe a não honrar com os compromissos de pagamento assumidos. Em alguns modelos, a pontuação pode ser literalmente interpretada como probabilidade de inadimplência, enquanto em outros, os pontos formam um sistema de classificação ao designar um tomador em um grupo de bons ou de maus

pagadores com base em um ponto limite (SAUNDERS, 2000, p. 13). Em Caouette, Altman e Naraynan (1999), Saunders (2000) e Crouhy, Galai e Mark (2004), são encontradas análises completas e as mais importantes metodologias utilizadas.

No entanto, somente em 1958, Bill Fair e Earl Isaac introduziram o primeiro modelo de *credit scoring* para a instituição financeira American Investments. Com o aumento da demanda após a criação dos cartões de crédito, na década de 60, e o aprimoramento dos recursos computacionais e tecnológicos, a implantação de modelos de *scoring* se tornou possível e passou a ser adotado em diferentes setores (IGUT, 2005).

Brigham (2000 apud QUEIROZ, 2006) argumenta que os modelos de *credit scoring* permitem, a partir de variáveis importantes e significativas, classificar e diferenciar clientes de crédito potencialmente bons de clientes potencialmente maus pagadores. Essa classificação é expressa em um valor numérico, facilitando a análise de muitos clientes de forma objetiva e imparcial. É largamente utilizado por bancos, financeiras, empresas de cartões de crédito, e mais recentemente, por grandes varejistas, empresas de telefonia, lojas de departamentos, entre outras.

A grande maioria dos modelos de *credit scoring* atribui pesos estatisticamente predeterminados as variáveis preditoras de inadimplência. Caso o *score* seja favorável, quando comparado ao valor de corte, então o crédito é aprovado. De um modo geral, são utilizadas técnicas estatísticas como análise discriminante e de regressão para o desenvolvimento desse tipo de modelo (CAOQUETTE; ALTMAN; NARAYNAN, 1999, p.180).

Os modelos baseados na análise discriminante múltipla predominaram até os anos 80, tendo como referência o modelo conhecido como Zeta, desenvolvido por Altman, Haldeman e Naraynan em 1977. O primeiro trabalho publicado com o emprego da regressão logística, também conhecida por *logit*, foi desenvolvido por Ohlson em 1980. A regressão logística também foi utilizada em vários estudos desde a década de 80, a exemplo da pesquisa de Wilson, Summers e Hope (2001) que incorporam dados sobre histórico de pagamento de 7.034 empresas, sendo 3.133 insolventes. Os autores concluíram que o comportamento pretérito de

pagamentos pode ser usado para prever futuros pagamentos dentro do contexto de crédito (EIFERT, 2003).

Saunders (2000) classifica os modelos de *credit scoring* em dois grandes grupos: modelos de aprovação de crédito e modelos de *score* comportamental, também conhecidos por *behavioural scoring*. Nos modelos de *behavioural scoring*, os dados históricos de pagamento e outras características já são conhecidos, de modo que é possível avaliar o comportamento de pagamento com base em operações anteriores. Nos modelos de aprovação de crédito, se pressupõe que o tomador ainda não possui um histórico de operações com a empresa.

Para Thomas (2000) há dois tipos de decisão que as empresas que emprestam ao consumidor têm que tomar. O primeiro está relacionada à concessão de crédito a um novo consumidor e utiliza ferramentas de *credit scoring* para auxiliar na tomada de decisão. O segundo tipo de decisão é de como tratar os clientes existentes. Caso um cliente queira aumentar o seu limite de crédito, a empresa deve concordar? Caso o cliente comece a atrasar os pagamentos, que ações devem ser tomadas pela empresa? Qual a ação de marketing adequada àquele tipo de consumidor? As técnicas que ajudam com essas tomadas de decisão são chamadas *behavioural scoring*. As informações adicionais inseridas nos modelos de *behavioural scoring* se referem ao histórico de compras e pagamentos realizados pelo consumidor.

Desta forma, enquanto o principal objetivo dos modelos de aprovação de crédito é estimar a probabilidade de um novo solicitante de crédito se tornar inadimplente com a instituição em determinado período, os modelos de *behavioural scoring* objetivam estimar a probabilidade de inadimplência de um cliente que já possui um produto ou crédito com a instituição.

Em uma concessionária de serviço público de distribuição de energia elétrica, o crédito é concedido antecipadamente, na forma de uma fatura de cobrança de consumos anteriores. Portanto, os modelos de *credit scoring* ou *behavioural scoring* constituem ferramentas de auxílio na otimização das operações de cobrança, tendo em vista restrições financeiras, regulatórias e operacionais. O

cálculo do desempenho futuro da carteira de cobrança, o monitoramento da carteira de consumidores e a melhoria do processo de cobrança são os principais benefícios advindos do desenvolvimento de modelos de atribuição de *score* de crédito. Portanto, verifica-se a possibilidade de utilização de técnicas estatísticas com o objetivo de discriminar bons e maus pagadores pode constituir uma importante ferramenta de auxílio na gestão da carteira de consumidores da DEE e ao planejamento das operações de cobrança realizadas em grande escala em uma empresa de grande porte.

4 TÉCNICAS DE ANÁLISE MULTIVARIADA UTILIZADAS

4.1 ANÁLISE MULTIVARIADA DE DADOS: DEFINIÇÕES E CONCEITOS GERAIS

Neste capítulo, objetiva-se apresentar a utilidade das técnicas de análise multivariada de dados usando exemplos ilustrativos.

A estatística multivariada compreende um conjunto de técnicas que analisam simultaneamente um conjunto de variáveis que caracterizam os objetos ou indivíduos de uma amostra. Usualmente as técnicas de estatística multivariada são classificadas em técnicas de dependência ou interdependência. Nas técnicas de dependência uma variável (variável dependente) é explicada por outras variáveis (variáveis independentes). Fazem parte das técnicas de dependência os modelos de regressão múltipla, regressão logística e a análise discriminante. Por sua vez, nas técnicas de interdependência nenhuma variável é considerada dependente ou independente, mas todas as variáveis são analisadas simultaneamente com a finalidade de encontrar uma estrutura para todo conjunto de variáveis. Fazem parte das técnicas de interdependência a análise fatorial, a análise de cluster e a análise de correspondência.

As técnicas de análise de dados, ou métodos de estatística descritiva multidimensional, segundo Lebart e outros (1985 apud TRIVELONNI; HOCHHEIM, 1999), classificam-se em duas grandes famílias: os métodos fatoriais e os métodos de classificação. Os métodos fatoriais utilizam cálculos de ajuste que recorrem essencialmente à álgebra linear e produzem representações gráficas onde os objetos a descrever se transformam em pontos sobre uma reta ou em um plano. Os métodos de classificação produzem classes ou famílias de classes que permitem agrupar e ordenar os objetos a descrever. Estas duas famílias de métodos são mais complementares que concorrentes e podem ser utilizadas proveitosamente de forma simultânea sobre um mesmo conjunto de dados.

Neste texto são descritas algumas técnicas de análise de agrupamentos (*cluster analysis*), mais precisamente métodos hierárquicos e não hierárquicos, assim como de análise de regressão logística (*logit*). Neste estudo, essas técnicas

são usadas na construção de um modelo para avaliação da inadimplência de um conjunto de consumidores de energia elétrica.

4.1.1 A variável estatística

A pesquisa científica só começa objetivamente depois da escolha das variáveis e das metodologias de análise, antes disto à atividade científica é completamente subjetiva. Ressalta-se que sempre existirá um viés subjetivo que interfere no resultado da pesquisa, haja vista que a análise estatística ocorre sobre as variáveis selecionadas pelo pesquisador (MOITA NETO, 2004).

De acordo com Hair e outros (2005), a variável estatística possui conceitos adicionais e questões relevantes que interferem na sua compreensão de aspectos como tipos de escalas de medida empregadas e estatísticas de teste de significância e níveis de confiança.

A variável estatística é uma combinação linear de variáveis com pesos empiricamente determinados pela técnica multivariada com um objetivo definido (HAIR e outros, 2005).

Uma variável estatística de n variáveis ponderadas (X_1 até X_n) pode ser enunciada matematicamente como:

$$\text{Valor da variável estatística} = w_1X_1 + w_2X_2 + w_3X_3 + \dots + w_nX_n$$

onde X_n é a variável observada e w_n é o peso determinado pela técnica multivariada.

4.1.2 Escalas de medida

Quantificar ou qualificar a variável é fundamental para possibilitar a descrição precisa do objeto pesquisado e a determinação da técnica estatística multivariada adequada para análise dos dados. Conforme assevera Hair e outros (2005), a medida de variação representa posição da variável em um conjunto de variáveis, seja entre elas ou entre uma variável dependente e uma ou mais variáveis independentes.

Os tipos de escala de medida que mais condicionam a análise estatística são as medidas contínuas ou métricas e as medidas categóricas ou não-métricas. Apesar de haver várias formas de classificar uma variável, neste trabalho foram adotadas apenas dois tipos de variáveis: métricas e não-métricas.

Uma variável métrica identifica ou descreve um objeto não apenas por identificar a existência do atributo, como também pela quantia ou grau que o objeto pode ser quantificado pelo atributo. Essa variável pode assumir qualquer valor em um intervalo, a exemplo de peso, corrente elétrica e altura. Já a variável não-métrica busca apenas identificar a presença ou ausência de atributos ou propriedades categóricas sem, no entanto, indicar a quantia. Por exemplo, a religião, o segmento de mercado e o grupo sanguíneo são dados não-métricos.

As variáveis não-métricas podem ser medidas através de uma escala nominal ou ordinal. As escalas nominais simplesmente representam categorias ou classes sem indicar a quantia. Nesse caso, os números ou símbolos utilizados para identificar os objetos não possuem significado quantitativo, apenas indicam a existência ou não da característica ou propriedade sob investigação. A utilização de uma escala ordinal possibilita que as variáveis não-métricas sejam ordenadas ou classificadas em um *ranking*, com base na quantia possuída pelo atributo, a exemplo dos diferentes níveis de satisfação do consumidor em relação aos diversos atributos da relação comercial com uma empresa (HAIR e outros 2005, p.24-28).

Por conveniência da análise, frequentemente é necessário criar novas variáveis com algumas categorias distintas que representem faixas de valores obtidas de um grande número de variáveis métricas (SPSS, 2005, p. 99). Através desse método, as variáveis métricas são transformadas em variáveis não-métricas que representam categorias. Por exemplo, a variável métrica “idade” pode ser transformada em uma nova variável não-métrica “faixa etária”.

4.1.3 Erro de medida e medida multivariada

Escalas de respostas inapropriadas, erros na entrada dos dados ou engano dos respondentes são causa da falibilidade ou imprecisão na medição dos

“verdadeiros” valores das variáveis. Segundo Hair e outros (2005, p.28), erro de medida é o grau em que os valores observados não são representativos dos valores “verdadeiros”. Pelo exposto, todas as variáveis estão sujeitas ao erro e podem acrescentar distorções às variáveis.

Para reduzir os efeitos do erro de medida, avalia-se a validade e a confiabilidade da medida. A validade refere-se à fidelidade da medida em relação ao que foi planejado, estando livre de qualquer erro sistemático ou não-aleatório. Por exemplo, se foi planejado obter a satisfação do responsável pelo domicílio, não se deve perguntar ao qualquer morador do domicílio. Nesse tipo de caso, o rigor no planejamento, na elaboração de questionário e no treinamento dos envolvidos torna-se importante para eliminar na origem os erros relacionados à validade. Por seu turno, a confiabilidade está relacionada com a consistência da variável ou conjunto de variáveis com o que se planejou medir. Refere-se ao modo como foi medido e não ao que foi medido. Cabe ao pesquisador avaliar as variáveis empregadas, sua consistência e existência de medidas alternativas para assegurar a maior confiabilidade possível.

Sobre o erro de medida, Hair e outros (2005, p. 29), explica que o impacto do erro de medida e a confiabilidade ruim dos mesmos não podem ser diretamente percebidos, uma vez que estão embutidos nas variáveis observadas. Indicando a necessidade de buscar uma descrição mais fiel das variáveis na direção de melhorar ou fortalecer os resultados da pesquisa.

4.1.4 Significância estatística e Poder estatístico

Ao investigar uma população através das variáveis de uma amostra, o pesquisador obriga-se a fazer inferências estatísticas sobre a amostra com o objetivo de obter informações para auxiliar no planejamento do estudo e na interpretação dos resultados. Com exceção de análise de agrupamentos e escalonamento multidimensional, todas as técnicas multivariadas são baseadas na inferência estatística, ou seja, nas relações entre a amostra e a população (HAIR e outros, 2005, p.29-30).

A significância estatística determina se o resultado da pesquisa pode ser atribuído ao acaso. Nesse sentido, faz-se necessário especificar os níveis de erro estatístico aceitáveis para possibilitar a interpretação dos dados. Hair e outros (2005) citam que a abordagem mais comum é especificar o erro estatístico admissível, denominado erro tipo I ou alfa (α). O erro Tipo I é a probabilidade de rejeitar incorretamente a hipótese nula. Significa dizer que existe uma diferença ou correlação quando na verdade não é o caso. Em suma, mede a chance de o teste exibir significância estatística sem que esta não esteja presente. Níveis comuns são cinco e um por cento, chamados de nível 0,05 ou 0,01, respectivamente.

Outro tipo de erro associado é o erro Tipo II ou beta (β). O erro Tipo II representa a probabilidade de não encontrar uma correlação ou diferença na média quando ela existe. Possui uma relação inversa com o erro Tipo I.

O poder de uma amostra, definido como valor $1 - \beta$ (erro Tipo II), é a probabilidade de não rejeitar a hipótese nula sendo ela falsa. Em outras palavras, poder é a probabilidade de encontrar corretamente um suposto relacionamento quando ele existe. O motivo mais frequente de erro tipo II é a pequena dimensão da amostra. Portanto, poder é a probabilidade de a significação estatística ser indicada se estiver presente. O poder é determinado por dois fatores, detalhados em Hair e outros (2005, p. 30) e resumidos a seguir.

a) Tamanho da amostra. Em qualquer nível alfa (erro tipo I), o aumento do tamanho da amostra provoca um aumento do poder do teste estatístico. Porém, ao elevar o tamanho da amostra, também pode aumentar o poder estatístico para níveis elevados.

b) Tamanho do efeito. Estimativa do grau em que o fenômeno estudado existe na população. Por exemplo, correlação ou diferença entre médias. Para avaliar o poder de qualquer teste estatístico, deve-se primeiro compreender o efeito que está sendo examinado. A métrica do efeito é padronizada. Desse modo, as diferenças de média são dadas em termos de desvios-padrão. Por exemplo, um

tamanho de efeito de 0,5 indica que a diferença de média é metade de um desvio-padrão.

4.2 EXAME DOS DADOS

Antes da aplicação de qualquer técnica multivariada, é prudente examinar e explorar a natureza dos dados e as relações entre variáveis. Busca-se desse modo, garantir que os métodos multivariados sejam aplicados em situações apropriadas e possibilitem interpretações mais corretas e esclarecedoras (HAIR e outros, 2005, p. 49-87).

O exame de dados consiste em uma etapa importante para superar possíveis problemas resultantes do planejamento da pesquisa e da coleta de dados. Segundo Hair e outros (2005, p.87),

a sofisticação analítica necessária para garantir que as exigências estatísticas sejam atendidas tem forçado o pesquisador a empregar uma série de técnicas de exame de dados que, em muitos casos, combinam com a complexidade das técnicas multivariadas.

Assim, a compreensão da natureza das variáveis e de suas relações possibilita um meio alternativo para desenvolver uma perspectiva sobre os dados especificamente sob um ponto de vista multivariado. As técnicas de exame de dados mais difundidas podem ser resumidas em i) análise de dados perdidos, ii) análise de observações atípicas e iii) teste das suposições da análise multivariada.

i) A técnica de exame de dados denominada “análise de dados perdidos” objetiva evitar que os dados perdidos afetem a generalização dos resultados. O dado perdido é causado por eventos como erros na entrada de dados, problemas na coleta de dados e recusa a responder o questionário. Essa técnica consiste em identificar a informação não disponível de uma variável que possui outras informações disponíveis.

ii) A observação atípica é uma observação que é substancialmente diferente das outras, ou seja, valores extremos em uma amostra ou população estudada. As observações atípicas devem ser identificadas no processo de tratamento e limpeza dos dados. Possuem duas causas. A primeira são os erros de procedimento, a

exemplo de erro na entrada de dados ou de falhas na codificação, que devem ser eliminadas ou registradas como valores perdidos. A segunda é resultado de um evento extraordinário, que explica o aspecto peculiar da observação. Cabe ao pesquisador decidir se o evento extraordinário deve permanecer na amostra.

A identificação de observações atípicas pode ser efetuada através de uma perspectiva univariada, bivariada e multivariada. A questão principal é a definição do conceito de observação atípica para a situação estudada. Na detecção sob a perspectiva univariada, o mais usual é converter os valores dos dados em escores padrão, que tem uma média de 0 (zero) e um desvio padrão de 1 (um). Quando as amostras são pequenas (80 observações ou menos), Hair e outros (2005) recomendam selecionar como atípicas as observações com escores padrão iguais ou maiores que 2,5. Para amostras maiores, recomenda-se como observações atípicas aquelas com escores padrão de 3 (três) a 4 (quatro).

A retenção ou eliminação de observações atípicas identificadas cabe ao pesquisador. Por diretriz, a menos que exista uma evidência de que as observações não representam a população, elas devem ser mantidas para garantir a generalidade à população como um todo.

iii) Teste das suposições estatísticas. A grande maioria dos testes estatísticos é baseada na distribuição normal diretamente ou em distribuições a ela relacionadas, como as distribuições t, F ou qui-quadrado. Tipicamente, estes testes requerem que as variáveis analisadas sejam normalmente distribuídas na população, ou seja, que elas atendam à "suposição de normalidade". Muitas variáveis observadas realmente são normalmente distribuídas, o que é outra razão por que a distribuição normal representa uma "característica geral" da realidade empírica. (STATSOFT INC, 2006).

A garantia de que para grandes amostras várias medidas estatísticas multivariadas têm distribuição próxima da normal, independente da forma da distribuição da população de onde foi extraída a amostra, é fundamentada no teorema do limite central. É bem conhecido o fato de que os valores entre ± 1 desvio padrão da média e ± 2 desvios padrões da média representam respectivamente 68,3% e 95,4% das observações. Medidas estatísticas, como média

e desvio padrão, são baseadas na teoria de distribuição normal, ou seja, na suposição que os valores das variáveis são distribuídos simetricamente em torno da média.

A multicolinearidade representa o grau em que uma variável pode ser explicada pelas outras variáveis na análise. Aumentos na multicolinearidade dificultam a análise da variável estatística e a determinação do efeito de qualquer variável individual devido ao alto grau de inter-relações existentes. Em análise de agrupamentos, as variáveis multicolineares são implicitamente ponderadas com maior peso.

Homocedasticidade é a variância constante dos resíduos. Esta é uma propriedade fundamental, que deve ser garantida, sob pena de invalidar toda a análise estatística. Deseja-se que os erros sejam aleatórios, ou seja, não devem ser relacionados com as características dos objetos estudados. Se isto não ocorre, há heterocedasticidade. Significa dizer que as chances de ocorrerem erros grandes (ou pequenos) variam conforme o tipo de objeto.

4.3 ANÁLISE DE AGRUPAMENTOS (CLUSTER ANALYSIS)

A análise de cluster é uma técnica de estatística multivariada que objetiva resolver o problema de como dividir um conjunto de N objetos, caracterizados por p atributos, em subconjuntos mutuamente exclusivos, denominados conglomerados ou clusters, de tal forma que os objetos em um mesmo subconjunto sejam semelhantes entre si, mas ao mesmo tempo diferentes dos objetos pertencentes aos outros subconjuntos. Problemas deste tipo são comuns em estudos de segmentação de mercado, na análise de portfólios e em estudos de *benchmarking* de empresas (TRIVELLONI; HOCHHEIM, 1999).

Em uma carteira de consumidores, a segmentação tem a finalidade de identificar grupos de consumidores com características semelhantes e levantar comportamentos típicos que podem ser encontrados na população. Esta informação é vital, pois através dela será possível identificar qual ou quais são os

grupos de consumidores e qual a melhor estratégia de minimização do risco de inadimplência e dos custos com ações de cobrança.

O termo análise de agrupamentos (*cluster analysis*) refere-se a uma variedade de algoritmos utilizados para classificar dados em estruturas que possibilitem inferir sobre as classes que apresentem características distintas entre si. Essa técnica é utilizada quando se deseja explorar as similaridades entre os indivíduos ou entre as variáveis de uma população.

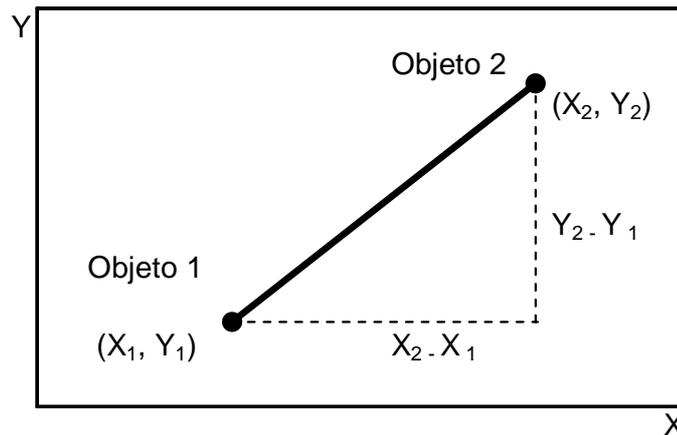
4.3.1 Medidas de similaridade

Todos os algoritmos de análise de clusters são baseados em uma medida de distância que procuram expressar o grau de semelhança entre os objetos. Em Hair e outros (2005), podem ser encontradas diversas medidas de distância, e a escolha da medida adequada vai depender da natureza qualitativa ou quantitativa dos atributos que caracterizam os objetos. Em análise de agrupamentos, a medida mais comumente usada é a distância euclideana, que é detalhada a seguir.

O conceito de distância euclideana deve ser familiar para a maioria dos pesquisadores, pois corresponde a aplicação do teorema de Pitágoras. Sejam $X = (X_1, \dots, X_p)$ e $Y = (Y_1, \dots, Y_p)$ dois objetos caracterizados por p atributos, então a distância euclideana entre estes dois objetos é definida por:

$$Distância_{X,Y} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2} \quad (4.1)$$

Um exemplo de como a distância euclideana é obtida é mostrado na Figura 4, a seguir.



$$\text{Distância} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

Figura 4 - Exemplo de distância euclidiana entre dois objetos medidos sobre duas variáveis

Fonte: Hair e outros (2005)..

As medidas de similaridade baseadas em distância euclidiana são as mais utilizadas. A distância euclidiana quadrada, é utilizada para acelerar o tempo de processamento, além de ser a distância recomendada para os métodos de agrupamento centróide e Ward, que serão detalhados a seguir (HAIR e outros, 2005, p.394).

Existem diversas medidas de distância, além da distância euclidiana, distância euclidiana quadrada, Minkovski, Block e distância de Mahalanobis descritos em Hair e outros (2005). A escolha da medida adequada vai depender da natureza qualitativa ou quantitativa dos atributos que caracterizam os objetos.

4.3.1.1 Coeficientes binários ou dicotômicos

As medidas de similaridade apresentadas até esse ponto são para tratamento de variáveis quantitativas. No entanto, as variáveis dicotômicas (binárias) também podem ser introduzidas neste tipo de análise à custa da sua transformação em variáveis binárias, com o valor 1 nos casos da presença de uma determinada característica de interesse e 0 nos casos contrários.

Os coeficientes binários expressam similaridades em termos de equiparações (*matches*) quando são utilizados dados qualitativos. Neste caso,

comparações são feitas entre observações reduzidas a duas categorias do tipo sim-não, presente-ausente, etc.

Muitas vezes os objetos não podem ser mensurados em variáveis quantitativas. Essas variáveis podem ser transformadas em dicotômicas (binárias), determinando um ponto de corte de interesse prático. Assim, por exemplo, se a altura (Y) de n indivíduos é mensurada e o interesse é determinar àqueles com altura superior a 1,90m, então, defini-se a variável binária (X) da seguinte forma: se $Y_i > 1,90m$ então $X_i = 1$ caso contrário, se $Y_i \leq 1,90m$, então $X_i = 0$. Da mesma forma, variáveis qualitativas podem ser transformadas em variáveis binárias tomando-se como valor 1 a presença de uma determinada realização e o valor 0 para as demais. Assim, por exemplo, se na amostra ocorresse um indivíduo com cor de olhos pretos determinaria o valor 1 e a ocorrência de outro com outra cor de olhos determinaria o valor 0. De uma maneira geral, a presença e ausência de uma característica devem ser representadas por uma variável binária, a qual assume valor 1 se a característica estiver presente e o valor zero se estiver ausente (FERREIRA, 1996, p.290) .

Para introduzir estas medidas de similaridade são apresentados na Figura 5 os resultados de coincidências e divergências dos objetos h e i em uma tabela de contingência. Na Tabela (apresentada na Figura 5) pode-se observar que “a” representa a frequência de coincidências (1-1), “b” a frequência de (1-0), e assim sucessivamente.

Indivíduo i	Indivíduo j		Totais
	1	0	
1	a	b	a + b
0	c	d	c + d
Totais	a + c	b + d	

Figura 5 - Regra de tratamento de variáveis dicotômicas (binárias)

Nota: Elaboração própria, adaptado de Ferreira (1996).

Conseqüentemente, a diferença ao quadrado entre os dois indivíduos ou objetos para uma determinada variável resultará apenas no valor 0 ou no valor 1.

Isso pode ser observado facilmente pelos seguintes argumentos.

$$(X_{hj} - X_{ij})^2 = \begin{cases} 0 & \text{se } X_{hj} = X_{ij} = 1 \text{ ou se } X_{hj} = X_{ij} = 0 \end{cases} \quad (4.2)$$

$$1 \text{ se } X_{hj} \neq X_{ij}$$

Dessa forma, a distância euclideana quadrática representa a contagem do número de pares não coincidentes. Grandes distâncias correspondem a muitos pares não coincidentes e, portanto, a objetos dissimilares.

A distância euclideana pode ser usada muitas vezes como base para cálculo da distância, no entanto, algumas vezes possui algumas limitações por considerar que os pares (1-1) e (0-0) possuem o mesmo peso, o que em determinadas situações reais (1-1) representa uma forte evidência de similaridade, mas o (0-0) não. Muitos coeficientes existem na literatura, dando diferentes tratamentos a este problema.

Cabe ao pesquisador decidir em qual situação o seu problema se enquadra e escolher a medida de parença mais apropriada.

Nesse estudo, os dados dicotômicos são considerados de igual peso atribuído às presenças e ausências simultâneas. Foi utilizada a “coincidência simples” (*simple matching coefficient*), como coeficiente de parença. A equação que explica os pesos iguais para 1-1 e 0-0 é $(a + d)/n$, que expressam a similaridade entre dois objetos em termos de equiparações positivas, isto é, presença de um mesmo caráter em ambos os objetos, em relação a equiparações negativas, isto é, ausência de um mesmo caráter em ambos os objetos, mais equiparações desencontradas. Portanto, esse algoritmo atribui pesos iguais tanto a presença, quanto a ausência da característica pesquisada.

4.3.2 Métodos de agrupamento

A classificação mais comum para representar os métodos de análise de agrupamentos apresenta dois grupos distintos. O primeiro é o método hierárquico, no qual os objetos são particionados em seqüência até atingir n agrupamentos, obtendo no final uma estrutura em árvore. O segundo é o método não-hierárquico, método pelo qual o conjunto é particionado em um número fixado de k agrupamentos.

4.3.2.1 Procedimentos hierárquicos de agrupamento

Nos métodos hierárquicos, os agrupamentos são construídos através de um processo iterativo que envolve uma combinação ou divisão de objetos em agrupamentos. Os dois procedimentos utilizados são os métodos aglomerativos e divisivos. No método divisivo, no início há apenas um agrupamento formado pelo conjunto de objetos que é dividido sucessivamente até que no final cada agrupamento contenha apenas um objeto. Enquanto no método aglomerativo, no início cada objeto forma um agrupamento que sucessivamente sofre uma série de fusões com outros agrupamentos até que no final todos os objetos estejam em um único agrupamento. Portanto, esse procedimento visa à unificação de objetos em blocos sucessivamente maiores. Uma ilustração dos dois métodos é apresentada na

Figura 6, a seguir.

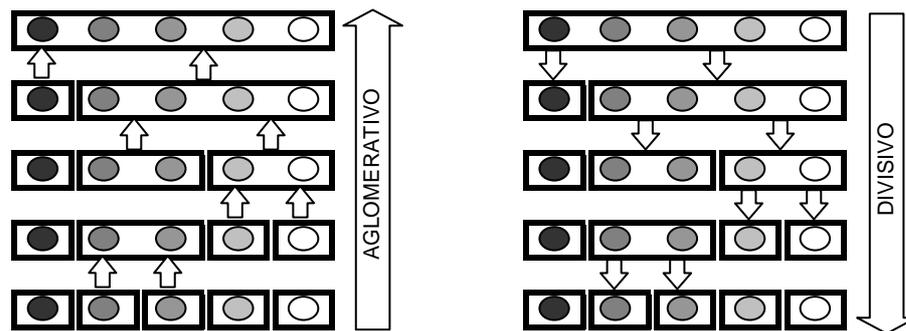


Figura 6 - Métodos aglomerativo e divisivo

Fonte: Trivelloni e Hochheim (1999)..

Para atingir este objetivo, utiliza-se alguma medida de similaridade ou de distância. Um resultado típico deste procedimento é uma árvore hierárquica, denominada dendrograma, que representa graficamente a formação dos agrupamentos. Um exemplo de dendrograma é apresentado abaixo.

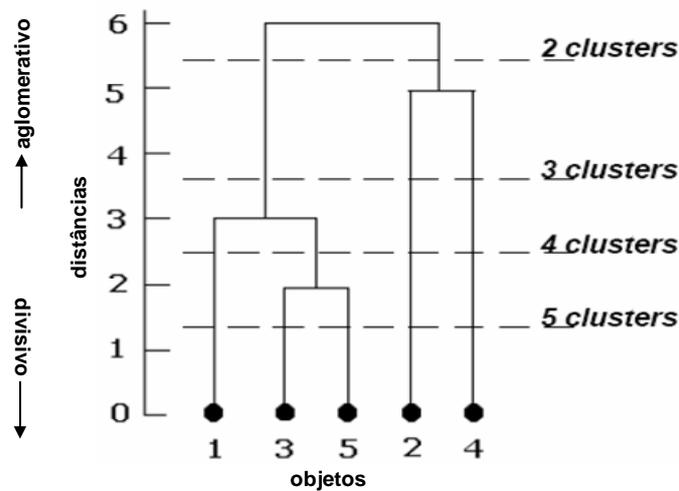


Figura 7 - Dendrograma ilustrativo do processo de agrupamento hierárquico

Fonte: Trivelloni e Hochheim (1999)..

Como resultado, tal processo obtém-se $n-1$ soluções de agrupamento, onde n é o número de objetos. Neste trabalho, utilizaram-se apenas métodos aglomerativos por serem mais comumente disponíveis na maioria dos pacotes estatísticos para computador além da constatação óbvia de que métodos divisivos são praticamente métodos aglomerativos ao contrário (HAIR e outros, 2005, p.398).

A aplicação dos procedimentos de agrupamentos impõe a necessidade de estimar as distâncias entre as classes já formadas e outras classes ou objetos. Os cinco algoritmos aglomerativos mais empregados são i) ligação individual (*single linkage method* ou *nearest neighbor*), ii) ligação completa (*complete linkage method* ou *farthest neighbor*), iii) ligação média (*average linkage* ou *median clustering*), iv) método de Ward (*Ward's method* ou *minimum variance clustering*) e v) método centróide (*centroid clustering*). Esses procedimentos diferem entre si na forma como a distância entre os agrupamentos é calculada. Não existe uma regra para definição do procedimento mais adequado. Recomenda-se que o pesquisador utilize vários algoritmos e avalie os resultados.

O método de ligação individual é baseado na distância mínima. Dados dois grupos (i,j) e (k), a distância entre eles é igual à menor distância entre os elementos dos dois grupos.

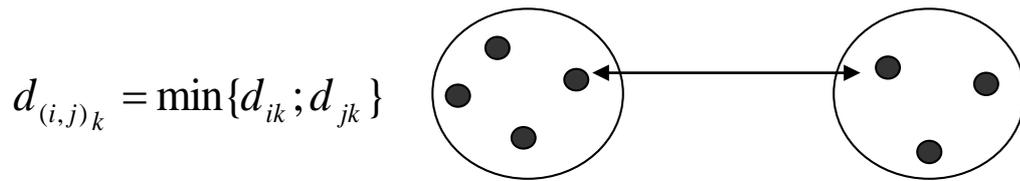


Figura 8 - Algoritmo ligação individual (single linkage method ou nearest neighbor).
Nota: Adaptado de Ferreira (1996).

Com este critério, cada indivíduo terá mais tendência para se agrupar a um grupo já definido do que para formar o núcleo de um novo grupo. Isto constitui uma desvantagem, principal responsável pela fraca utilização deste critério.

No método de ligação completa, a distância entre o grupo formado e outro é igual à maior das distâncias entre os elementos dos dois grupos. Dados dois grupos (i,j) e (k), a distância entre eles é igual à maior distância entre os elementos dos dois grupos.

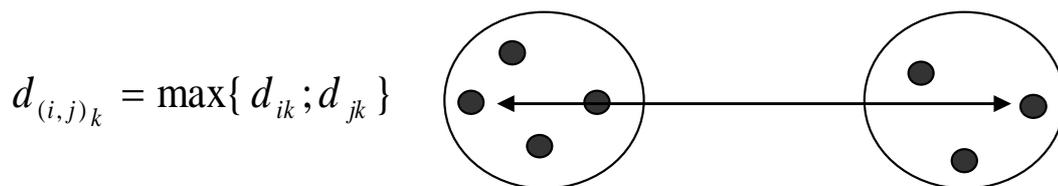


Figura 9 - Algoritmo ligação completa (complete linkage method ou farthest neighbor).

Nota: Adaptado de Ferreira (1996).

Com este critério, cada grupo passa a ser definido como o conjunto dos indivíduos em que cada um é mais semelhante a todos os outros do grupo do que a qualquer outro elemento. Os grupos assim criados são mais compactos.

No método de ligação média, a similaridade é representada através da distância média entre todos os objetos em um agrupamento e todos os objetos de outro agrupamento. Dados dois grupos (i,j) e (k), a distância entre eles é a média entre todos os pares de indivíduos constituídos por todos os elementos dos dois grupos. De acordo com Hair e outros (2005), essa abordagem tende a produzir agrupamentos com aproximadamente a mesma variância.

O método de Ward utiliza o critério de menor erro, ou perda de informação, considerando o somatório dos quadrados das diferenças entre cada elemento ou cluster a ser agrupado e o valor central de cada cluster já formado. Para identificar

no conjunto de elementos existentes qual a seqüência de agrupamento a ser adotada é avaliado o aumento deste erro para cada combinação dos clusters existentes.

O erro é definido como:

$$SQE = \sum_{i=1}^n (X_i - \bar{X})^2 \quad (4.3)$$

Este critério é baseado na comparação entre a aplicação da medida da soma dos quadrados dos desvios das observações em relação à média dos grupos. Portanto, a distância entre dois agrupamentos é a soma dos quadrados entre os dois agrupamentos feita sobre todas as variáveis. Primeiro são calculadas as médias das variáveis de cada grupo, em seguida, é calculado o quadrado da distância euclideana entre essas médias e os valores das variáveis para cada indivíduo. Por fim, somam-se as distâncias para todos os indivíduos e otimiza-se a variância mínima dentro dos grupos (HAIR e outros, 2005, p.401).

No método centróide, a similaridade entre agrupamentos é medida como a distância entre centróides de agrupamento. Dados dois grupos, a distância entre eles é a distância entre os respectivos centróides, média das variáveis caracterizadoras dos indivíduos de cada grupo. Em Hair e outros (2005), está descrito que os centróides são alterados a cada fusão de objetos, que podem produzir resultados confusos. Esse método hierárquico é o menos afetado por observações atípicas.

A seguir é apresentado o algoritmo do método aglomerativo para formar clusters em um conjunto com n objetos:

- a) Inicie com n clusters, cada um contendo apenas um objeto e construa a matriz de distâncias de ordem n ;
- b) Identifique o menor elemento da matriz de distâncias para encontrar o par de clusters mais similares;

- c) Reúna os dois clusters identificados na etapa 2 em um único cluster e atualize a matriz de distâncias, retirando as linhas e colunas relativas aos dois clusters identificados em 2 e incluindo a linha e coluna com as distâncias entre os demais clusters e o novo cluster formado. Note que a ordem da matriz de distâncias diminui de uma unidade a cada vez que a etapa 3 é executada;
- d) Repita os passos 2 e 3 até que reste apenas um cluster. A cada iteração guarde a identificação dos clusters que foram fundidos e também a distância entre eles, estas informações serão utilizadas na montagem do diagrama de árvore ou dendrograma que mostra a seqüência de aglomeração dos clusters.

Como não há critério estatístico para a determinação no número ideal de agrupamentos (regra de parada), recomenda-se que o pesquisador adote pelo menos uma regra de parada que é relativamente simples. Primeiro examinar o dendrograma resultante do agrupamento e verificar visualmente a ocorrência de grandes elevações na medida de distância. Segundo, examinar a matriz de similaridade ou distâncias entre os agrupamentos em cada passo sucessivo. A definição da quantidade ideal de agrupamentos ou clusters pode ser percebida quando há um salto repentino nos valores da distância interna média.

4.3.2.2 Procedimentos não-hierárquicos de agrupamento

Os procedimentos não-hierárquicos produzem apenas uma solução de agrupamentos para um conjunto de sementes de agrupamento. Diferentemente do processo de construção de um dendrograma utilizado no método hierárquico, as sementes de agrupamento são empregadas para reunir objetos dentro de uma distância pré-especificada das sementes. Nos procedimentos não-hierárquicos não são produzidos resultados para todos os possíveis números de agrupamentos, logo não é necessária a construção da matriz de similaridade, cuja ordem depende do número de objetos. Quando se trata de amostras muito grandes, as exigências de armazenamento de dados aumentam fortemente, de maneira que os métodos não-

hierárquicos são computacionalmente mais viáveis quando objetiva-se formar agrupamentos em um conjunto com uma grande quantidade de objetos. Os métodos de agrupamentos não-hierárquicos são populares através da aplicação do agrupamento de k médias ou *K-Means* (HAIR e outros, 2005, p.402-403).

Os agrupamentos não-hierárquicos objetivam determinar o agrupamento de n objetos em k grupos. Nesses métodos o número de clusters k é previamente fixado, ao contrário do que ocorre nos métodos hierárquicos, onde a definição do número de clusters derivada da análise do dendrograma. Por esta razão, recomenda-se testar diversos valores de k para minimizar o risco de aceitar agrupamentos duvidosos.

O algoritmo k -médias, de uma forma bastante simplificada, é dividido em três passos. Neste método, a separação de um conjunto de n objetos em k clusters é feita segundo o algoritmo abaixo:

- a) Escolha um conjunto de k centróides iniciais. Os centróides são pontos quaisquer no espaço p -dimensional e a escolha dos centróides iniciais pode ser aleatória ou baseada em algum outro critério. O emprego de pontos sementes não-aleatórios assegura as vantagens sobre o método hierárquico, a exemplo da menor susceptibilidade às observações atípicas.
- b) Percorra a lista de objetos e aloque cada objeto ao centróide mais próximo. Note que após alocar todos os objetos, foram formados agrupamentos.
- c) A partir dos agrupamentos formados calcule novos centróides para todos os agrupamentos e volte para o passo 2. O algoritmo deve continuar até o momento que não haja realocação de objetos de um agrupamento para outro.

Um exemplo do funcionamento do algoritmo do agrupamento não-hierárquico é apresentado a seguir. Considere um conjunto com 4 objetos ($n=4$), com as coordenadas na tabela abaixo, onde se procura formar dois grupos ($k=2$).

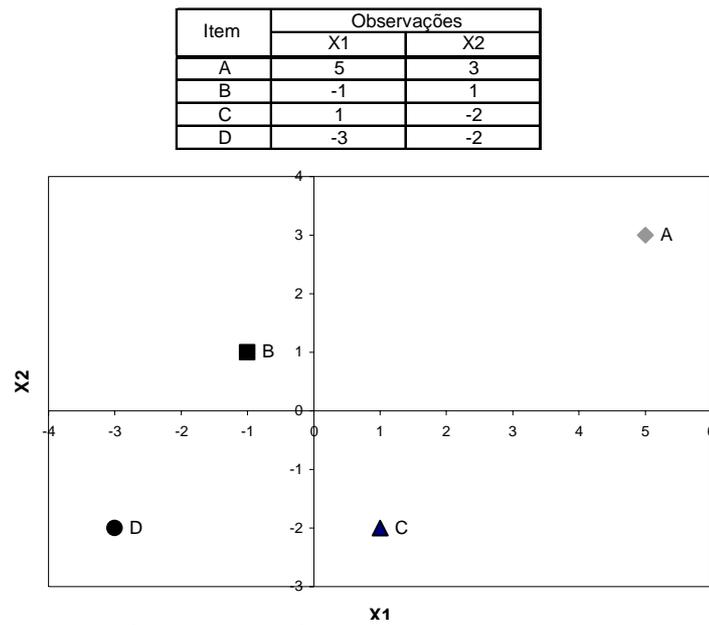


Figura 10 - Configuração dos quatro objetos para o agrupamento não-hierárquico
Fonte: Trivelloni e Hochheim (1999)..

Optou-se por formar aleatoriamente dois grupos AB e CD e a partir deles definir as coordenadas dos centróides.

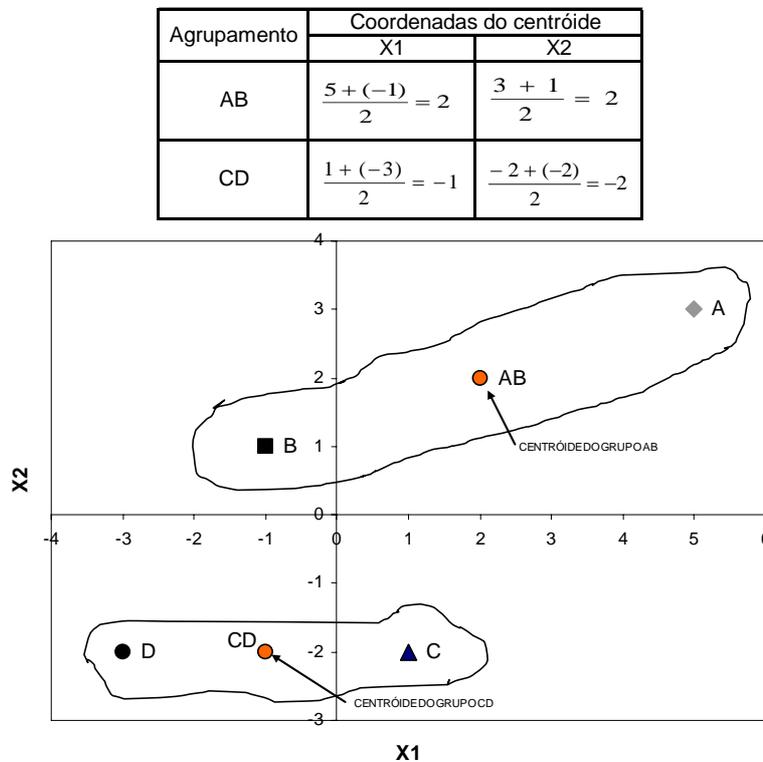


Figura 11 - Grupos AB e CD com os respectivos centróides
Fonte: Trivelloni e Hochheim (1999)..

Uma vez determinadas as coordenadas dos centróides, são calculadas as distâncias de cada objeto com relação aos centróides para verificar a necessidade de realocação. Iniciando com o objeto A, têm-se as seguintes distâncias ao quadrado:

$$D^2(A, (AB)) = (5 - 2)^2 + (3 - 2)^2 = 10 \quad (4.4)$$

$$D^2(A, (CD)) = (5 + 1)^2 + (1 + 2)^2 = 45 \quad (4.5)$$

Como a distância entre A e o centróide AB é menor que a distância entre A e o centróide CD, não há realocação. Assim, A permanece no grupo AB. Seguindo agora com o objeto B têm-se as seguintes distâncias ao quadrado:

$$D^2(B, (AB)) = (-1 - 2)^2 + (1 - 2)^2 = 10 \quad (4.5)$$

$$D^2(B, (CD)) = (-1 + 1)^2 + (1 + 2)^2 = 9 \quad (4.6)$$

Neste caso a distância entre B e o centróide CD é a menor distância, então o objeto B deve ser realocado para o grupo CD e as novas coordenadas dos centróides são:

Agrupamento	Coordenadas do centróide	
	X1	X2
A	5	3
BCD	-1	-1

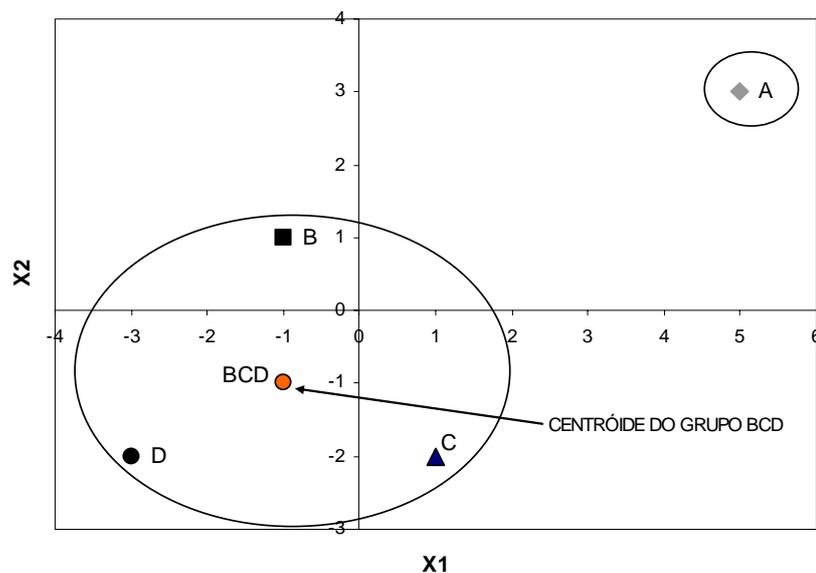


Figura 12 - Grupos A e BCD com o respectivo centróide
Fonte: Trivelloni e Hochheim (1999)..

Continuando com os objetos C e D verifica-se que não há realocação.

A tabela abaixo (apresentada na Figura 13) mostra as distâncias ao quadrado de cada objeto com relação aos centróides. Note que em todos os casos o quadrado da distância do objeto ao centróide do grupo onde está alocado é sempre a menor distância, assim não há realocação de objetos e a execução do algoritmo pode ser finalizada.

Agrupamento	Distâncias ao quadrado			
	Ítem			
	A	B	C	D
A	0	40	41	41
BCD	52	4	5	5

Figura 13 - Distâncias ao quadrado dos Grupos A e BCD

Fonte: Trivelloni e Hochheim (1999)..

Mesmo com o emprego de uma solução inicial não-aleatória, não há a garantia de que um agrupamento ótimo foi formado. Assim, recomenda-se testar diferentes valores de k .

Conforme recomenda Hair e outros (2005), neste trabalho optou-se por empregar uma combinação dos dois métodos com o objetivo de alcançar os benefícios de ambos. Primeiro, é realizada uma classificação hierárquica pelo método de Ward, buscando-se determinar o número de agrupamentos, caracterizar os centros de agrupamentos e identificar quaisquer observações atípicas. Segundo, é realizado o agrupamento das observações pelo método *k-means* adotando como pontos sementes iniciais os indicados na análise hierárquica.

4.3.3 Exemplo de uma análise de agrupamentos

A análise de agrupamentos pode ser ilustrada por um simples exemplo. Parte-se da suposição de que o interesse é determinar segmentos de consumidores em uma dada localidade, com base em padrões de percepção da qualidade da fatura mensal de energia elétrica. Uma pequena amostra de sete respondentes é selecionada para apresentar resumidamente como a análise de agrupamentos é

aplicada. Cinco variáveis são consideradas para cada respondente: V₁ (entrega da conta de luz com antecedência), V₂ (conta de luz sem erros, leitura do medidor e cálculos corretos), V₃ (facilidade de entendimento das informações que estão na conta), V₄ (locais de pagamento) e V₅ (dia do vencimento da conta de luz). As variáveis são apresentadas em uma escala de 0-10. Os valores da resposta de cada um dos sete respondentes, são mostrados na Tabela 4 e na Figura 14, que compreende um diagrama de perfil representativo das notas atribuídas pelos sete respondentes às cinco variáveis da pesquisa.

Tabela 4 - Valores de dados das sete observações com base nas cinco variáveis de agrupamento

Variável de Agrupamento:		Respondentes							
Características da Concessionária		a	b	c	d	e	f	g	h
V1	Entrega da conta de luz com antecedência	7,0	9,0	5,0	6,0	1,0	4,0	2,0	10,0
V2	Conta de luz sem erros, leitura do medidor e cálculos corretos	10,0	9,0	5,0	6,0	2,0	3,0	4,0	2,0
V3	Facilidade de entendimento das informações que estão na conta	9,0	8,0	6,0	3,0	2,0	2,0	5,0	4,0
V4	Locais de pagamento	7,0	9,0	7,0	3,0	1,0	3,0	2,0	2,0
V5	Dia do vencimento da conta de luz	10,0	9,0	7,0	4,0	2,0	3,0	5,0	6,0

Nota: Elaboração própria.

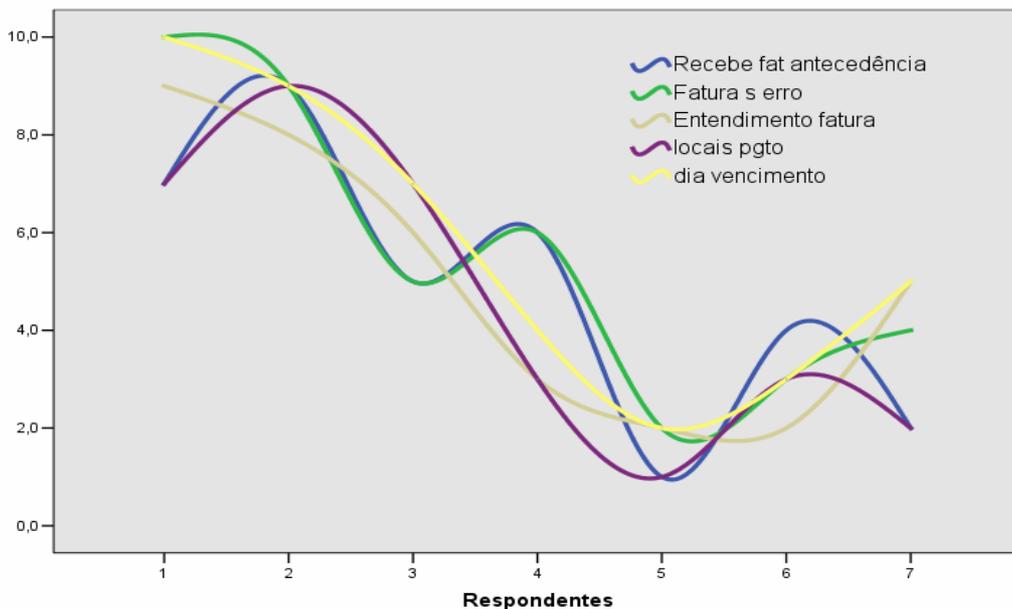


Figura 14 - Diagrama de perfil das sete observações com base nas cinco variáveis de agrupamentos

Nota: Elaboração própria.

A demonstração a seguir foi adaptada do roteiro apresentado em Hair e outros (2005).

4.3.4 Objetivo da análise de agrupamentos

O objetivo é segmentar uma população de consumidores em grupos com percepções semelhantes sobre a fatura mensal de energia elétrica. Portanto, busca-se obter tanto uma classificação de objetos baseada na experiência do pesquisador, retratada pela seleção das variáveis. Uma vez agregados os consumidores com base em características a eles associadas, torna-se possível desenvolver estratégias específicas, com diferentes apelos, para cada um dos agrupamentos.

Em qualquer aplicação de análise de agrupamentos, a seleção das variáveis deve considerar o conhecimento sobre quais as variáveis relevantes para o entendimento do fenômeno analisado. Para o exemplo ilustrativo apresentado, assume-se que os atributos selecionados possuem poder preditivo capaz de revelar as relações entre as observações.

4.3.5 Projeto de pesquisa em análise de agrupamentos

Uma vez definidos os objetivos da pesquisa e selecionadas as variáveis, deve-se verificar na amostra a existência de observações atípicas e, se for o caso, avaliar se devem ser eliminadas. O processo de partição da população também deve ser precedido da definição sobre como a similaridade dos objetos deve ser medida e se os dados devem ser normalizados. O resultado da aplicação da técnica de análise de agrupamentos é impactado por decisões do pesquisador em cada estágio. Conforme Hair e outros (2005, p. 391), “as questões do projeto de pesquisa e da escolha de metodologias feita pelo pesquisador talvez tenham maior impacto em análise agrupamentos do que com qualquer outra técnica multivariada”. Nessa etapa, se faz o exame dos dados, a exemplo da análise de dados perdidos e de observações atípicas e dos testes das suposições da análise multivariada.

Para o exemplo descrito no item 4.3.5, a avaliação da amostra de sete observações com o objetivo de identificar dados atípicos, não encontrou um forte candidato para eliminação. No entanto, em estágios posteriores, deve-se avaliar se observações atípicas surgiram durante o processo de agrupamento. A similaridade será medida de acordo com a distância euclidiana entre cada par de observações, por ser a mais simples e intuitiva. Distâncias menores indicam maior similaridade entre observações.

Na Tabela 5 é apresentada a matriz de proximidade contendo as distâncias euclidianas entre as observações. Os dados utilizados medem a magnitude das percepções dos consumidores dentro de uma escala de notas entre zero e dez. Por esse motivo, os dados apresentam corretamente a intensidade da percepção dos consumidores, não sendo necessário normalizar as variáveis.

Tabela 5 - Matriz de proximidades de distâncias euclidianas entre observações

Observação	Euclidean Distance						
	a	b	c	d	e	f	g
a	-						
b	3,317	-					
c	6,856	6,633	-				
d	10,247	10,198	6,000	-			
e	15,780	16,186	10,100	7,071	-		
f	13,115	13,000	7,280	3,873	3,873	-	
g	11,269	12,166	6,325	5,099	4,899	4,359	-

Fonte: Output do SPSS.

4.3.6 Determinação de agrupamentos e avaliação do ajuste geral

Após o cálculo da matriz de proximidades, o processo de partição é iniciado com a definição do algoritmo para formar os agregados e decidir o número de agrupamentos a serem formados. Optou-se por ilustrar a aplicação combinada dos métodos hierárquicos e não-hierárquicos. Os métodos hierárquicos são empregados para especificar os pontos sementes de agrupamentos para aplicação no método não-hierárquico.

O primeiro passo é a aplicação do procedimento hierárquico aglomerativo. O método de Ward é adotado para formar os agrupamentos. Na Tabela 6, são detalhados os passos do processo hierárquico aglomerativo.

No passo 1, as observações com menor proximidade entre si (a-b) são alocadas em um agrupamento, reduzindo de um total de sete para seis agrupamentos. No passo 2 são encontrados os próximos pares de observações mais semelhantes (e-f). No passo 3, o agrupamento (e-g) com (e-f) são combinados formando um novo agrupamento (e-f-g). No passo 4 o agrupamento (d-e) é combinado com o agrupamento (e-f-g), formado o agrupamento (d-e-f-g). Até esse passo, de um total de sete agrupamentos, restaram apenas três agrupamentos. No passo 5 o par (a-c) é combinado com o par (a-b) formando o agrupamento (a-b-c). O passo final consiste na combinação dos dois agrupamentos remanescentes formando um único agrupamento.

Tabela 6 - Processo de agrupamento hierárquico aglomerativo para as sete observações

Ward's method											
Passo	Processo de aglomeração		Solução de agrupamento			Coeficiente de aglomeração					
	Distância mínima entre observações não-agrupadas ^a	Par de observações	Pertinência a agrupamento		Número de agrupamentos						
	Solução inicial		(a)	(b)	(c)	(d)	(e)	(f)	(g)		
1	3,317	a-b	(a-b)	(c)	(d)	(e)	(f)	(g)		6	1,658
2	3,873	e-f	(a-b)	(c)	(d)	(e-f)	(g)			5	3,595
3	4,899	e-g	(a-b)	(c)	(d)	(e-f-g)				4	6,035
4	7,071	d-e	(a-b)	(c)	(d-e-f-g)					3	8,952
5	6,856	a-c	(a-b-c)	(d-e-f-g)						2	12,895
6	10,247	a-d	(a-b-c-d-e-f-g)							1	25,378

^a Distância euclideana entre observações.

Nota: Elaboração própria.

A representação gráfica do processo de agrupamento é demonstrada no dendrograma da Figura 15. O dendrograma permite uma inspeção visual em busca de observações atípicas e como instrumento de apoio à decisão do número de agrupamentos na solução final. Nesse caso, as possíveis soluções variam de um a seis agrupamentos. Sabe-se que quando se afasta dos agrupamentos unitários a homogeneidade diminui. No entanto, uma solução com sete agrupamentos não permite a definição de qualquer estrutura subjacente aos dados. Quando se avança para o passo 6, ou seja, para a solução de apenas um agrupamento, percebe-se um grande aumento na distância euclideana, indicando que a junção desses dois últimos agrupamentos resulta em um agregado bem menos homogêneo. Portanto, tanto a variação do coeficiente de aglomeração demonstrado na Tabela 6, quanto a verificação manual do dendrograma apresentado na Figura 15, sugere que a solução com dois agrupamentos parece ser mais adequada para uma solução final por reduzir a quantidade de agrupamentos e assegurar a diferenciação entre eles.

4.3.7 Interpretação e validação do perfil dos agrupamentos

A solução de dois agrupamentos resultou em grupos com o mesmo tamanho (3 e 4 observações) e em um perfil correspondente ao do procedimento hierárquico. O primeiro agrupamento é composto dos respondentes que percebem um serviço prestado com boa qualidade, enquanto no segundo ficaram os respondentes que não percebem uma boa qualidade do serviço. Nenhuma variável apresentou diferença significativa entre os agrupamentos. Portanto, a correspondência e estabilidade entre os procedimentos hierárquicos e não-hierárquicos validam a decisão de adotar dois agrupamentos.

Obteve-se agrupamentos de uma população de consumidores em grupos com percepções semelhantes sobre a fatura mensal de energia elétrica. Uma vez agregados os consumidores com base em características a eles associadas, torna-se possível desenvolver estratégias específicas, com diferentes apelos, para cada um dos agrupamentos.

4.4 ANÁLISE DE REGRESSÃO LOGÍSTICA

A partir de Ohlson (1980) a Regressão Logística ou modelo LOGIT tem sido usado frequentemente para a avaliação de riscos de inadimplência.

Segundo Hair e outros (2005), a Regressão Logística é uma forma especial de regressão na qual a variável dependente é não-métrica, dicotômica, binária. Apesar de algumas diferenças, a maneira geral de interpretação é semelhante à da regressão linear.

Os coeficientes logísticos da regressão podem ser usados para estimar relações da probabilidade da variável dependente com cada uma das variáveis independentes utilizadas no modelo (SPSS INC, 2004, p.13).

A Regressão Logística (RL) e a Análise Discriminante são as técnicas mais adequadas quando a variável é categórica (nominal ou não-métrica) e as variáveis independentes são métricas. A RL possui a vantagem de ser menos afetada quando as suposições estatísticas, especificamente de normalidade das variáveis não são

satisfeitas. A RL também pode acomodar variáveis independentes não-métricas por meio da transformação em variáveis dicotômicas (HAIR e outros, 2005, p.210). A aplicação em uma escala mais ampla de situações pode ser explicada porque o modelo de regressão logística requer bem menos suposições que a análise discriminante e, mesmo quando as suposições requeridas para a análise discriminante são satisfeitas, a regressão logística continua apresentando bons resultados (BOGDANSKI; TOMBINI; WERLANG, 2000).

A aplicação da RL pode ser exemplificada com o caso de investigação das características do estilo de vida que são fatores de risco para a doença de coração *Coronary Heart Disease* (CHD). Dado uma amostra dos pacientes avaliados quanto a uso de fumo, dieta, exercício físico e uso do álcool, é possível construir um modelo usando as quatro variáveis do estilo de vida e predizer a presença ou a ausência de CHD em uma amostra dos pacientes. O modelo pode então ser usado para predizer a probabilidade de ocorrência da CHD, assim como identificar as relações das probabilidades para cada fator do estilo de vida, como, por exemplo, quanto mais fumantes provavelmente desenvolverão CHD do que não fumantes.

4.4.1 Estimação do modelo de Regressão Logística

A RL prediz diretamente a probabilidade de um evento ocorrer em uma escala de valores entre zero e um. A definição dos valores preditos é obtida através da relação entre a variável dependente e as variáveis independentes que forma uma curva logística em forma de “s”, também chamada de sigmóide. Uma função sigmóide é capaz de “imitar” uma série de comportamentos dependendo de sua inclinação, que é ajustada de acordo com os pesos das variáveis. Na Figura 16 é apresentada a forma da relação logística. No início da curva logística ocorre um crescimento a taxas crescentes, que caso persistisse indefinidamente levaria o nível da variável independente ao infinito. A partir de certo momento, o crescimento passa a evoluir a taxas decrescentes até atingir um “ponto de saturação” (HAIR e outros, 2005, p.232).

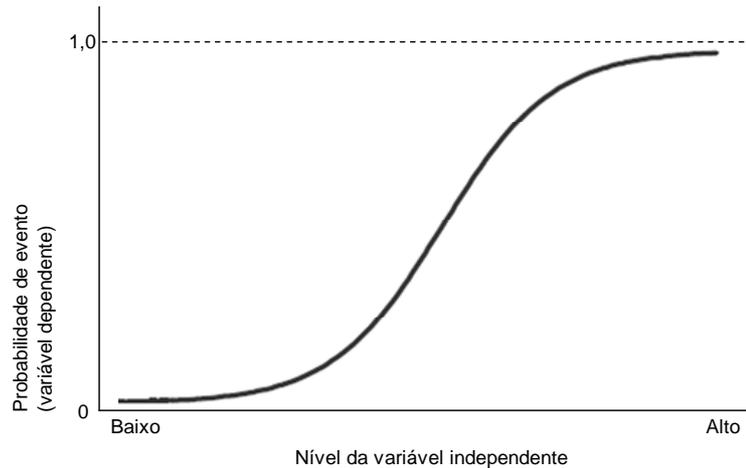


Figura 16 - Forma da relação logística entre variáveis dependente e independente

Fonte: HAIR e outros (2005).

A Figura 17 representa dois exemplos hipotéticos de ajuste de uma relação logística aos dados da amostra, extraídos de Hair e outros (2005). Os dados reais, ou seja, a ocorrência ou não de um evento (0 ou 1) em função de cada valor da variável independente, são representados como observações no topo ou na base do gráfico. Na parte “A”, há uma clara sobreposição dos eventos não sendo possível obter um bom ajuste dos dados. Em contrapartida, na parte “B” não há sobreposição das distribuições apresentando uma relação bem mais definida com dados bem ajustados ao modelo.

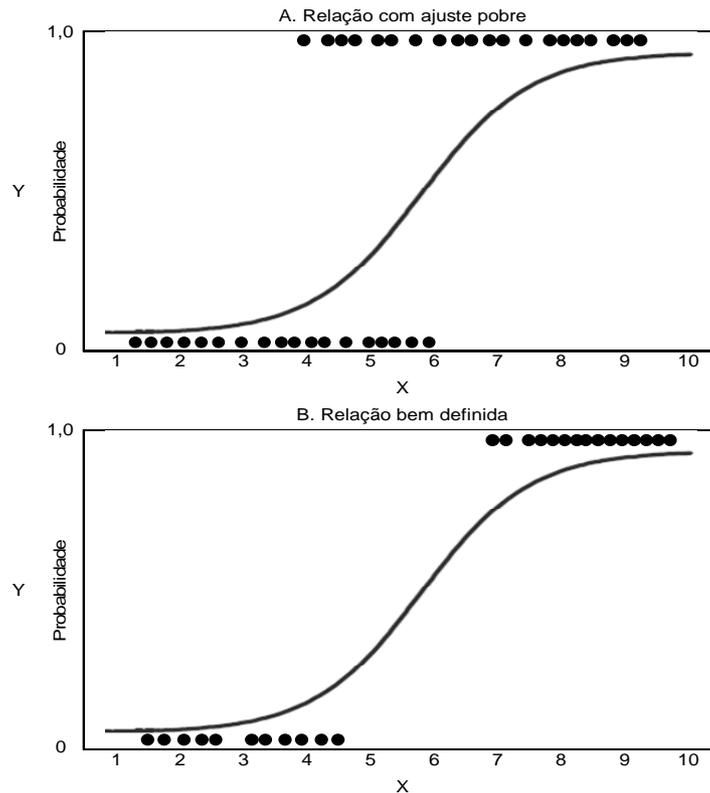


Figura 17 - Dois exemplos hipotéticos de ajuste de uma relação logística aos dados de uma amostra
 Fonte: Hair e outros (2005).

No contexto da prevenção de inadimplência de consumidores de energia elétrica, a variável resposta compreende um comportamento de pagamento, definido por um conjunto de informações disponíveis na base de dados da concessionária de distribuição de energia elétrica. Nesse sentido, espera-se melhorar a capacidade de avaliação da carteira de consumidores e priorizar ações que reduzam o risco de inadimplência.

A regressão logística (*logistic regression* ou *logit analysis*) é uma abordagem de modelagem matemática usada para descrever a relação entre diversas variáveis independentes e uma variável dependente dicotômica (HAIR e outros, 2005).

O modelo logístico é baseado na função logística $f(z)$, em forma de “S”, dada pela equação a seguir:

$$f(z) = \frac{1}{1 + e^{-(z)}} \quad (4.7)$$

onde,

$e \approx 2,718$;

Z é uma combinação linear das variáveis independentes: $Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_pX_p$;

B_0 uma constante e $B_1 \dots B_p$ coeficientes estimados a partir dos dados;

$X_1 \dots X_p$ as variáveis independentes.

O modelo de regressão logística é determinado pela seguinte relação:

$$\log(\pi)/(1 - \pi) = \beta_0 + \beta_1 x_1 + \beta_p x_p$$

Inicia-se a análise considerando a existência de uma variável contínua Z, não observada, que pode dar uma idéia da “propensão para” a ocorrência do evento em estudo. Por exemplo, em uma carteira de consumidores, Z representa a propensão de um consumidor inadimplir no pagamento da conta mensal pelo consumo de energia elétrica. Valores maiores de Z correspondem a altas probabilidades de inadimplência (SPSS INC, 2004).

4.4.2 Interpretação dos coeficientes

De acordo com Hair e outros (2005, p. 233), uma das vantagens da regressão logística é simplicidade da variável resposta, uma vez que se necessita apenas saber se o evento ocorreu para então usar um valor dicotômico como uma variável dependente. Por exemplo, risco de crédito ou não, compra ou não, falência de empresa ou sucesso. A partir desse valor dicotômico, o método é capaz de prever a probabilidade de ocorrência ou não do evento.

A comparação da probabilidade de um evento com a probabilidade de o evento não acontecer é denominada razão de desigualdade, a qual é usada como variável dependente em regressão logística. A razão dessa desigualdade pode ser expressa através da seguinte equação:

$$\frac{\text{Probabilidade}_{(\text{evento})}}{\text{Probabilidade}_{(\text{evento})}} = e^{B_0 + B_1X_1 + \dots + B_nX_n} \quad (4.8)$$

Os coeficientes estimados B_0, B_1, \dots, B_n representam medidas de variação na relação das probabilidades, que é chamada razão de desigualdade. Eles são expressos na forma de logaritmos e podem ser transformados de volta para que seus efeitos relativos sobre as probabilidades sejam avaliados com maior facilidade. Desse modo, um coeficiente positivo aumenta a probabilidade, enquanto que um coeficiente negativo diminui a probabilidade prevista (HAIR e outros, 2005, p.233).

4.4.3 Avaliação da adequação do ajuste do modelo e teste da significância dos coeficientes

Com o objetivo de verificar a significância do modelo final, existem dois testes possíveis (HAIR e outros, 2005, p.234). O primeiro deles é o teste do qui-quadrado, que se baseia no desvio entre o modelo ajustado e o que se chama de modelo saturado, ou seja, um modelo onde n parâmetros ajustam perfeitamente n observações. Quanto menor for o desvio, medido pela estatística $-2LL$ ($-2 \log\text{-likelihood}$), melhor é o modelo. O outro teste conduzido é o teste de Hosmer-Lemeshow. Esse teste indica se há ou não diferença estatística entre as classificações esperadas e as observadas no modelo.

O teste do qui-quadrado, apresenta a medida de melhoria do modelo devido a introdução de cada variável independente. É obtido através de uma medida chamada "pseudo R^2 ". Para um modelo *logit* (R^2_{logit}) pode ser calculado através da equação seguinte.

$$R^2_{logit} = \frac{-2LL_{nulo} - (-2LL_{modelo})}{-2LL_{nulo}} \quad (4.9)$$

A outra forma de validar o modelo é o teste de Hosmer-Lemeshow. Este teste avalia o modelo ajustado comparando as freqüências observadas e as esperadas. O teste associa os dados as suas probabilidades estimadas da mais baixa a mais alta, então faz um teste qui-quadrado para determinar se as freqüências observadas estão próximas das freqüências observadas. A ausência de significação estatística indica um bom ajuste do modelo. (HAIR e outros, 2005).

A análise de Regressão Logística tem sido amplamente utilizada em modelos de previsão de insolvência por indicar diretamente a probabilidade e por não exigir que a distribuição das variáveis seja normal. Press e Wilson (1978) e Krzanowaski (1988) consideram que há praticamente um consenso de que a Regressão Logística deve ser preferida quando as distribuições claramente não atendem à suposição de normalidade. Portanto, o fato de não necessitarmos da suposição de normalidade multivariada, torna a Regressão Logística mais genérica.

Neste capítulo, foi apresentada uma visão geral da Regressão Logística, técnica que será utilizada neste trabalho para construção de um modelo de previsão de inadimplência de consumidores de energia elétrica.

5 ESTUDO DE CASO

Nesse capítulo, apresenta-se a aplicação da metodologia para avaliação do risco de inadimplência da carteira de recebíveis hipotética, construída a partir dos dados da Coelba, com o objetivo de identificar as características que discriminam a melhor representação da estrutura dos consumidores e possibilitar a adoção de um conjunto de estratégias de cobrança para minimizar a inadimplência e os custos dela decorrentes. Nesse estudo, desenvolvem-se a classificação da carteira de consumidores em agrupamentos homogêneos, a análise do comportamento de pagamento e o cálculo de equações de regressão que explicam a formação da perda esperada ou probabilidade de inadimplência.

Inicialmente, apresenta-se a caracterização da carteira de consumidores com base no mês de junho de 2006 e o plano amostral utilizado no estudo. A formação dos grupos de consumidores privados por atividade econômica (CNAE) é realizada através da rotina *complex samples* do SPSS. O nível de confiança para os estratos de consumidores por CNAE é de 97%, com uma margem máxima de erro de 5%.

5.1 SEGMENTAÇÃO DA CARTEIRA DE CONSUMIDORES

5.1.1 Objetivos da segmentação e análise dos dados dos consumidores inadimplentes

O objetivo é estabelecer uma política de cobrança baseada em segmentos de consumidores privados de energia elétrica disponíveis. Foi utilizado como base as variáveis listadas nas Tabelas 8 e 9, consideradas relevantes e adequadas para obtenção de segmentos que sejam caracterizados pelo consumo médio e atividade econômica. Portanto, busca-se obter uma classificação de objetos baseada na experiência do pesquisador, retratada pela seleção das variáveis. A premissa é que através da associação entre atividade econômica, consumo por classe e faixa de consumo dos consumidores privados inadimplentes é possível obter segmentos de consumidores que propiciem o desenvolvimento de estratégias específicas de cobrança. O estudo é constituído com base na carteira de consumidores existentes na base do “contas a receber” do fechamento contábil de junho de 2006. A carteira

é constituída de 2.752.422 consumidores ligados que possuíam pelo menos uma fatura vencida ou a vencer.

Tabela 8 - Distribuição do número de consumidores com débito por classe e faixa de consumo

Consumidores privados com débito por faixa de consumo e CNAE *							
Atividade econômica	Unid.	0-50 kWh	50,01-300 kWh	300,01-500 kWh	500,01-1000 kWh	Acima 1000 kWh	TOTAL
RESIDENCIAL	qtd	827.146	1.497.022	41.242	12.889	3.290	2.381.589
	%	34,7%	62,9%	1,7%	0,5%	0,1%	100,0%
COMERCIAL	qtd	55.758	88.089	15.917	13.089	13.444	186.297
	%	29,9%	47,3%	8,5%	7,0%	7,2%	100,0%
INDUSTRIAL	qtd	2.061	5.932	1.427	1.479	2.899	13.798
	%	14,9%	43,0%	10,3%	10,7%	21,0%	100,0%
RURAL	qtd	78.115	76.823	5.869	5.003	4.925	170.735
	%	45,8%	45,0%	3,4%	2,9%	2,9%	100,0%
TOTAL	qtd	963.081	1.667.868	64.455	32.460	24.558	2.752.422
	% fx	35,0%	60,6%	2,3%	1,2%	0,9%	100,0%

Notas: CNAE - Cadastro Nacional de Atividade Econômica. Consumo médio mensal 12 meses
Elaboração própria.

Na Tabela 8 é apresentado o número de consumidores privados com débito, distribuídos por quatro classes de consumidores e cinco faixas de consumo. Observa-se que 95,6% dos consumidores inadimplentes situam-se nas faixas de consumo até 300 kWh. Essa grande concentração de consumidores em faixas de menor consumo representa uma elevada dispersão dos créditos em consumidores com faturas mensais de menor valor. Logo, o benefício x custo de cobrar esse tipo de consumidor sempre está na agenda de qualquer empresa, particularmente naquelas que possuem seus custos controlados pelo Agente Regulador.

Por outro lado, ao analisar na Tabela 9 a distribuição do consumo desses mesmos consumidores, em comparação com os dados da Tabela 8, observa-se um comportamento oposto. Enquanto os consumidores das faixas de consumo acima de 300 kWh respondem por apenas 4,4% dos consumidores privados inadimplentes, o volume de consumo associado responde por 61,5% de todo o consumo. Portanto, verifica-se a existência de um pequeno grupo de consumidores com maior susceptibilidade ao controle da inadimplência, justificando ações de cobrança de maior custo frente ao valor do faturamento médio por consumidor.

Tabela 9 - Distribuição do consumo de consumidores com débito por classe e faixa de consumo

Consumo faturado de consumidores privados com débito por faixa de consumo e CNAE *							
Atividade econômica	Unid.	0-50 kWh	50,01-300 kWh	300,01-500 kWh	500,01-1000 kWh	Acima 1000 kWh	TOTAL
RESIDENCIAL	kWh	22.612.856	148.878.069	15.332.319	8.370.953	6.741.129	201.935.326
	%	11,2%	73,7%	7,6%	4,1%	3,3%	100,0%
COMERCIAL	kWh	1.289.520	11.837.189	6.143.549	9.135.920	90.790.052	119.196.230
	%	1,1%	9,9%	5,2%	7,7%	76,2%	100,0%
INDUSTRIAL	kWh	48.567	867.069	549.187	1.028.055	135.631.451	138.124.330
	%	0,0%	0,6%	0,4%	0,7%	98,2%	100,0%
RURAL	kWh	1.750.810	7.986.244	2.259.625	3.467.640	46.045.205	61.509.524
	%	2,8%	13,0%	3,7%	5,6%	74,9%	100,0%
TOTAL	kWh	25.701.754	169.568.571	24.284.681	22.002.568	279.207.838	520.765.413
	%	4,9%	32,6%	4,7%	4,2%	53,6%	100,0%

Notas: * CNAE - Cadastro Nacional de Atividade Econômica. Consumo médio mensal 12 meses.

Elaboração própria.

A Coelba possui mais consumidores na faixa de consumo “50,01-300 kWh” em todas as classes, com forte predominância na classe Residencial, que agrega 89,8% desses consumidores. Essa mesma faixa responde por 32,6% do consumo estudado. A faixa de consumo “0-50 kWh” possui 85,9% dos consumidores na classe Residencial e responde por apenas 4,9% do consumo da população estudada. Na faixa “acima 1000 kWh” está 0,9% dos consumidores fortemente concentrados em 53,6% do total do consumo. Nessa faixa, o consumo concentra-se nas classes Industrial, Comercial e Rural. O número de consumidores nas três faixas situadas até 500 kWh corresponde a 97,9% do total e a 42,2% do consumo indicando uma importante concentração de 57,8% do consumo em 2,1% dos consumidores.

No exame do relacionamento entre faixa de consumo e atividade econômica, verifica-se que as faixas de consumo “acima de 300 kWh” se aproximam mais dos consumidores das atividades Industrial e Comercial. Para as faixas de consumo abaixo de 300 kWh, todas as atividades econômicas possuem uma expressiva participação.

Os dados apresentados mostram uma substancial variação entre as categorias de consumo por atividade econômica. A faixa de consumo “50,01-300 kWh” possui as mais altas concentrações de consumidores para todas as atividades. A quantidade de consumidores das classes Comercial e Industrial também é relevante na maior faixa “acima 1000 kWh”. Por outro lado, e como era de se esperar, a faixa “Acima 1000 kWh” possui menores concentrações de

consumidores das classes Residencial e Rural que concentram as maiores quantidades na faixa “50,01-300 kWh”. Em uma leitura por atividade econômica observa-se que 99,3% dos consumidores da atividade residencial situam-se nas faixas de consumo até 500 kWh, seguida das atividades rural e comercial com 94,2% e 85,8%, respectivamente. Em contrapartida, 31,7% dos consumidores da atividade industrial concentram-se na faixa de consumo acima de 500 kWh. De uma maneira geral, a distribuição dos 2.752.419 consumidores da amostra sugere uma grande variação no consumo por consumidor e a existência de um pequeno bloco de consumidores, equivalente a 2,1%, que é constituído basicamente de indústrias e estabelecimentos comerciais.

Assim, pode-se concluir que o consumo faturado dos consumidores com débito, medido em kWh, possui capacidade de discriminar segmentos de consumidores com a finalidade de estabelecer políticas de cobrança alinhadas com a importância relativa dos consumidores e com os objetivos da empresa.

O exame das Tabelas 8 e 9 sugere a existência de quatro agrupamentos de consumidores. O primeiro, de consumo elevado, acima de 1000 kWh, representado basicamente de consumidores das classes industrial e comercial. O segundo, de consumo médio, na faixa de 300,01 a 1000 kWh, composto basicamente de consumidores das atividades residencial, comercial, rural e industrial de pequeno porte. O terceiro, de pequeno consumo, de 50,01 a 300 kWh, com predominância de consumidores das atividades residencial, rural e comercial de pequeno porte. O quarto, de baixo consumo, até 50 kWh, composto em sua maioria por consumidores beneficiários da tarifa social baixa renda das classes residencial previsto na Resolução ANEEL n°. 246, de 30 de abril de 2002¹¹. Portanto, frente ao largo espectro de consumo dentro das mesmas classes de consumidores, considerou-se mais adequado a utilização de uma segmentação considerando o consumo médio como a principal variável discriminante.

Os resultados dos estudos reforçam a estruturação de quatro segmentos de consumidores privados, detalhados na Tabela 10, a seguir. Observa-se uma escala

¹¹ Estabelece as condições para enquadramento na subclasse residencial baixa renda da unidade consumidora com consumo mensal inferior a 80 kWh.

piramidal com uma forte inclinação, na qual poucos consumidores respondem por uma grande concentração de consumo. Por exemplo, aproximadamente 4,4% dos consumidores respondem por cerca de 63% do consumo total faturado.

Tabela 10 - Segmentos de consumidores privados de energia elétrica

Segmento	Sigla	Qtd. %	Consumo %	Características
1 - Privados Baixo Consumo	PCB	35,00%	4,90%	<i>Consumidores privados com consumo médio até 50 kWh.</i>
2 - Privados Consumo Pequeno	PCP	60,60%	32,60%	<i>Consumidores privados com consumo médio entre 50,01 e 300 kWh.</i>
3 - Privados Consumo Médio	PCM	3,50%	8,90%	<i>Consumidores privados com consumo médio entre 300,01 e 1000 kWh.</i>
4 - Privados Consumo Alto	PCA	0,90%	53,60%	<i>Consumidores privados com consumo médio acima de 1000 kWh.</i>

Nota: Elaboração própria.

O segmento PCB agrega consumidores privados com consumo médio até 50 kWh. Possui 35% dos consumidores inadimplentes e 4,9% do consumo. Esse segmento caracteriza-se pela elevada quantidade de unidades de baixo consumo beneficiadas pela tarifa social para consumidores de baixa renda. É formado em sua grande maioria por consumidores das atividades Residencial e Rural. O baixo valor por fatura de energia requer uma política de cobrança com ações de baixo custo e medidas preventivas de incentivo ao uso racional da energia.

O segmento PCP agrega consumidores privados com consumo médio entre 50,01 e 300 kWh, Esse segmento é caracterizado pela elevada quantidade de unidades de pequeno consumo e pequena densidade de kWh por consumidor. Possui uma distribuição equitativa de consumidores pertencentes a todas as atividades econômicas. A implantação da política de cobrança é impactada por restrições orçamentárias e operacionais que impõem o uso otimizado dos recursos na busca de minimizar o risco de inadimplência.

No segmento PCM, que agrega consumidores privados com consumo médio entre 300,01 e 1000 kWh, verifica-se uma distribuição mais uniforme dos consumidores por classe econômica com uma presença destacada das atividades Industrial e Comercial. Apenas a classe residencial é pouco expressiva. A

existência de uma maior densidade de kWh por consumidor justifica uma política de cobrança com menor restrição orçamentária e a aplicação efetiva de todos os meios de cobrança necessários à manutenção do adimplemento.

O segmento PCA é formado por consumidores privados com consumo médio acima de 1000 kWh. É predominantemente composto de consumidores da classe industrial com pequena presença das classes comercial e rural. Nesse segmento participação da classe residencial é inexpressiva. A política de cobrança requer um ciclo reduzido de dias e o monitoramento rigoroso do comportamento de pagamento, uma vez que é elevado risco de inadimplência.

5.2 ESTUDO DO COMPORTAMENTO DE PAGAMENTO

Após a definição da segmentação dos consumidores de energia elétrica pertencentes ao setor privado, segue-se avaliando as características disponíveis na base de dados da concessionária associadas ao comportamento de pagamento. O estudo de comportamento de pagamento é dividido em duas partes. A primeira analisa a formação de blocos de consumidores com base na atividade econômica. A segunda parte analisa a formação de agrupamentos com base nos segmentos de consumidores apresentados na Tabela 10.

5.2.1 Análise de agrupamentos com base na caracterização da dívida

Nessa análise, serão utilizadas as técnicas de análise hierárquica de agrupamentos e análise não-hierárquica com o objetivo de comparar os resultados obtidos e validar o modelo.

Objetiva-se então identificar o relacionamento entre consumidores com base em variáveis dicotômicas. Por exemplo, presença de dívida vencida e não paga, de dívida na Provisão com Créditos de Liquidação Duvidosa (PCLD), de débito em conta corrente, se é consumo baixa renda. Finalmente, os agrupamentos identificados serão associados às atividades econômicas Residencial, Residencial Baixa Renda, Comércio e Indústria. Uma vez identificados grupos com características semelhantes, a concessionária pode formular estratégias com

diferentes abordagens de cobrança para grupos específicos de consumidores. Para uma população finita, composta pelos contratos de consumidores privados, a amostra utilizada de 800 consumidores assegura um nível de confiança de 95% e uma margem máxima de erro desejada de 3,5%. A seleção da amostra foi obtida através da rotina *select samples* do SPSS. A amostra foi analisada com o objetivo de buscar dados atípicos, não sendo identificado casos que justifiquem uma eliminação de qualquer observação.

Nessa análise de agrupamento foram utilizadas cinco variáveis, relacionadas a seguir.

- 1) Possui fatura negativada no SPC/SERASA pela Coelba (sim = 1/não=0).
- 2) Possui fatura baixa renda (sim = 1/não=0).
- 3) Possui dívida vencida (sim = 1/não=0).
- 4) Possui dívida na PCLD (sim = 1/não=0).
- 5) Possui débito em conta-corrente (sim = 1/não=0).

O exame das cinco variáveis usadas na análise de agrupamentos não apresentou distorções que justificassem a redução dos dados ou recálculo das medidas de distâncias. Assim, não foram encontradas evidências da existência de multicolinearidade que afetasse a análise.

As variáveis quantitativas e qualitativas utilizadas foram transformadas em variáveis dicotômicas (binárias), as quais assumem valor 1 se a característica estiver presente e o valor zero se estiver ausente.

As exigências de normalidade, linearidade e homocedasticidade que de fato são importantes em outras técnicas, possuem pouco peso em análise de agrupamentos. No entanto, a representatividade da amostra e a multicolinearidade são analisadas visando assegurar uma adequada representação da estrutura dos dados.

5.2.1.1 Análise hierárquica de agrupamentos

O algoritmo de agrupamento utilizado foi o método centróide. Nesse método, a similaridade é medida através da distância euclidiana quadrada entre centróides de agrupamentos. Quando dois agregados são combinados, um novo centróide é calculado. Logo, os centróides de agrupamentos migram ou se movem conforme os agrupamentos são combinados. Essa técnica é menos afetada por observações atípicas com relação a outros métodos hierárquicos (HAIR e outros, 2005). Os resultados da análise de agrupamentos, incluindo os casos em combinação em cada estágio do processo e o coeficiente de aglomeração, são apresentados no ANEXO A. A quarta coluna representa o coeficiente de aglomeração obtido através da soma interna de quadrados entre os dois agrupamentos somados sobre todas as variáveis.

As cinco variáveis analisadas, listadas no item 5.2.1, envolvem aspectos relevantes quanto ao comportamento de pagamento dos consumidores da Coelba. Portanto, objetiva-se identificar tipos ou perfis de consumidores que possam formar as bases para diferentes estratégias de relacionamento e cobrança. A seguir, com base no esquema de aglomeração apresentado no ANEXO A e no dendrograma obtido através da análise hierárquica pelo método centróide apresentado no ANEXO B, será selecionada uma solução final de agrupamentos.

Segundo Hair e outros (2005), o coeficiente de aglomeração é particularmente útil para aplicação em uma regra de parada que avalie a evolução do coeficiente em cada estágio do processo de aglomeração pelo método hierárquico. Pequenos coeficientes indicam que agrupamentos bem homogêneos estão sendo fundidos. A junção de dois agregados muito diferentes resulta em um grande coeficiente ou uma grande variação percentual no coeficiente.

Para analisar as tendências da regra de parada e específica utilizada na análise, considerou-se um intervalo de soluções possíveis entre dois e dez agrupamentos. Na Tabela 11, são apresentados os 10 últimos estágios do processo de agrupamento hierárquico detalhado no ANEXO A, com o coeficiente de

aglomeração e a variação percentual no coeficiente de um para dois agrupamentos, e assim sucessivamente até o número de 10 agrupamentos.

O coeficiente de aglomeração sofre uma primeira grande alteração na passagem de dois para um agrupamento (-16,75%). A próxima grande variação ocorre na passagem de seis para cinco agrupamentos (-21,90%). Assim, duas soluções são consideradas para avaliação da significância prática do agrupamento, as soluções de 2 e 6 agregados. Portanto, buscar-se-á identificar se as soluções são úteis e substanciais o bastante para garantir uma ação prática.

Tabela 11 - Análise do coeficiente de aglomeração para análise hierárquica de agrupamentos pelo método centróide

Número de agrupamentos	Coeficiente de aglomeração	Variação Percentual no coeficiente para o próximo nível
10	-0,11468	-0,04
9	-0,11473	-15,30
8	-0,13229	-15,63
7	-0,15296	-3,06
6	-0,15764	-21,90
5	-0,19217	-0,06
4	-0,19205	-2,54
3	-0,19692	-1,24
2	-0,19937	-16,75
1	-0,23276	-

Nota: Elaboração própria, adaptado do output SPSS.

5.2.1.2 Identificação de observações atípicas

Através do dendrograma (ANEXO B) e do esquema de aglomeração (ANEXO B), ambos obtidos do pacote estatístico SPSS, é possível identificar observações atípicas, onde um dado desses seria um “ramo” que não se juntou com outro, ou se juntou muito tarde. Também é possível identificar pequenos agrupamentos, uma vez que eles exibem um “ramo” muito longo para apenas um pequeno número de observações.

No ANEXO A, as colunas ao lado direito do coeficiente apresentam o estágio em que o agrupamento surgiu pela primeira vez (*stage cluster first appears*) anotando os passos nos quais cada agrupamento foi formado. Uma observação que nunca foi reunida em um agrupamento tem o estágio zero. Até o estágio 398, as observações isoladas são colocadas juntas. Somente no estágio 399 ocorre a

primeira junção com um outro agrupamento formado em um estágio anterior. Não se localizou observações individuais candidatas a dados atípicos.

Conclui-se que os resultados obtidos através da regra de parada, ausência de dados atípicos e perfis distintos de consumidores são adequados para uma solução de dois e de seis agrupamentos. Portanto, a análise das duas soluções será levada adiante na análise não-hierárquica para “refinar” o estudo e auxiliar na decisão final.

5.2.1.3 Análise de agrupamentos não-hierárquica

Nesse estudo, os procedimentos não-hierárquicos visam validar os resultados obtidos na análise hierárquica. Os dados foram processados no SPSS através do método *K-means*, utilizando dois e seis agrupamentos sugeridos na análise hierárquica anterior. Os valores dos centróides e o tamanho dos agrupamentos para as soluções de dois e seis agrupamentos são apresentados na Tabela 12. Os perfis correspondem bem aos perfis do procedimento hierárquico. Observa-se que a solução de dois agrupamentos possui os mesmos valores de centróides. Nas duas soluções apenas um grupo absorveu a maioria dos consumidores presentes na amostra. Na solução de dois agrupamentos, um grupo concentra 81,5% da amostra, enquanto na solução de seis agrupamentos um grupo concentra 75,3% da amostra.

Tabela 12 - Perfis das variáveis com as soluções de dois e seis agregados da análise hierárquica de agrupamentos

Agrupamento	Valores médios					Tamanho do agrupamento
	V ₁	V ₂	V ₃	V ₄	V ₅	
	Possui fatura negativada na Coelba	Possui fatura Baixa renda	Possui dívida vencida	Possui dívida na PCLD	Possui débito em conta-corrente	
<i>Solução de dois agrupamentos - Centros do agrupamento final</i>						
1	0	0	1	0	0	148
2	0	0	1	0	0	652
					total	800
<i>Solução de seis agrupamentos - Centros do agrupamento final</i>						
1	0	0	1	0	0	602
2	0	0	0	0	1	27
3	0	1	1	0	0	44
4	1	0	1	0	0	56
5	0	0	1	1	0	46
6	1	1	0	0	0	25
					total	800

Nota: Elaboração própria, adaptado do output SPSS.

A Tabela 13 apresenta o teste de significância de diferenças entre centros de agrupamento para as soluções com dois e seis agrupamentos. Na solução de dois agrupamentos, com exceção da variável V5 (débito em conta corrente), todas as demais variáveis são estatisticamente significantes. Já na solução com seis agrupamentos todas as variáveis são estatisticamente significantes, fornecendo uma estrutura melhor definida e com maior diferenciação entre as variáveis.

Tabela 13 - Teste de significância de diferenças entre centros de agrupamento
Solução de dois agrupamentos

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
NEG_COEL	20,768	1	,054	798	384,451	,000
FAT_BX_RENDA	11,221	1	,062	798	181,509	,000
DÍVIDA VENCIDA?	26,255	1	,206	798	127,199	,000
DÍVIDA NA PDD?	25,463	1	,046	798	552,815	,000
DEBITO_EM_C_C ORRENTE	,094	1	,035	798	2,689	,101
<i>Solução de seis agrupamentos</i>						
	Mean Square	df	Mean Square	df	F	Sig.
NEG_COEL	10,737	5	,013	794	836,444	,000
FAT_BX_RENDA	10,343	5	,011	794	928,721	,000
DÍVIDA VENCIDA?	7,376	5	,194	794	38,007	,000
DÍVIDA NA PDD?	9,491	5	,019	794	510,330	,000
DEBITO_EM_C_C ORRENTE	5,198	5	,002	794	2105,153	,000

Fonte: Output do SPSS.

Conforme apresentado na Tabela 14, na solução de dois agrupamentos, a diferenciação ocorre porque o *cluster 1* absorveu os consumidores com dívida na PCLD. Também se verifica que a proporção de consumidores por atividade econômica entre o *cluster 1* e o *cluster 2* é uma característica diferenciadora. O *cluster 1* apresenta uma importante predominância de consumidores da atividade residencial (43,9%) e menor participação da atividade comercial (13,5%). Por outro lado, o *cluster 2* possui uma distribuição mais uniforme de consumidores por atividade econômica com apenas 19,2% da atividade residencial, 27,8% da atividade rural, que predomina ligeiramente em relação às outras.

Tabela 14 - Distribuição de consumidores por atividade econômica: 2 agrupamentos

		Cluster Number of Case		Total
		1	2	
Residencial	Count	65	125	190
	% within CNAE	34,2%	65,8%	100,0%
	% within Cluster Number of Case	43,9%	19,2%	23,8%
	% of Total	8,1%	15,6%	23,8%
Comercial	Count	20	180	200
	% within CNAE	10,0%	90,0%	100,0%
	% within Cluster Number of Case	13,5%	27,6%	25,0%
	% of Total	2,5%	22,5%	25,0%
Industrial	Count	31	166	197
	% within CNAE	15,7%	84,3%	100,0%
	% within Cluster Number of Case	20,9%	25,5%	24,6%
	% of Total	3,9%	20,8%	24,6%
Rural	Count	32	181	213
	% within CNAE	15,0%	85,0%	100,0%
	% within Cluster Number of Case	21,6%	27,8%	26,6%
	% of Total	4,0%	22,6%	26,6%
Total	Count	148	652	800
	% within CNAE	18,5%	81,5%	100,0%
	% within Cluster Number of Case	100,0%	100,0%	100,0%
	% of Total	18,5%	81,5%	100,0%

Fonte: Output do SPSS.

Conforme apresentado na Tabela 15, na solução de seis agrupamentos, o *cluster 1* absorveu a maior parcela de consumidores por cada atividade econômica: Residencial (55,8%), Comercial (84,0%), Industrial (79,2%) e Rural (80,8%). Esse agrupamento, juntamente com o *cluster 2*, caracteriza-se pela presença de consumidores bons pagadores com nenhuma dívida vencida ou, no máximo com

uma fatura vencida. A diferença entre os dois agregados de bons pagadores é que o *cluster 2* absorveu os consumidores com débito em conta-corrente - DCC. O *cluster 6* compreende os consumidores que estavam em transição para a adimplência e podem ser absorvidos pelo *cluster 1*, pois não possuem DCC. Os *clusters 3 e 4* absorveram os consumidores com dívida vencida, porém, ainda sem dívida provisionada na PCLD. O *cluster 3* circunscreve os consumidores da atividade Residencial baixa renda sem negativação. Os consumidores com maior risco de inadimplência foram selecionados no *cluster 5*. Por já possuírem dívidas vencidas, inclusive provisionadas à perdas, representam dívidas de difícil recebimento.

Tabela 15 - Distribuição de consumidores por atividade econômica: 6 agrupamentos

		Cluster Number of Case						Total
		1	2	3	4	5	6	
Residencial	Count	106	3	44	10	9	18	190
	% within CNAE	55,8%	1,6%	23,2%	5,3%	4,7%	9,5%	100,0%
	% within Cluster Number of Case	17,6%	11,1%	100,0%	17,9%	19,6%	72,0%	23,8%
	% of Total	13,3%	,4%	5,5%	1,3%	1,1%	2,3%	23,8%
Comercial	Count	168	9	0	8	12	3	200
	% within CNAE	84,0%	4,5%	,0%	4,0%	6,0%	1,5%	100,0%
	% within Cluster Number of Case	27,9%	33,3%	,0%	14,3%	26,1%	12,0%	25,0%
	% of Total	21,0%	1,1%	,0%	1,0%	1,5%	,4%	25,0%
Industrial	Count	156	6	0	21	10	4	197
	% within CNAE	79,2%	3,0%	,0%	10,7%	5,1%	2,0%	100,0%
	% within Cluster Number of Case	25,9%	22,2%	,0%	37,5%	21,7%	16,0%	24,6%
	% of Total	19,5%	,8%	,0%	2,6%	1,3%	,5%	24,6%
Rural	Count	172	9	0	17	15	0	213
	% within CNAE	80,8%	4,2%	,0%	8,0%	7,0%	,0%	100,0%
	% within Cluster Number of Case	28,6%	33,3%	,0%	30,4%	32,6%	,0%	26,6%
	% of Total	21,5%	1,1%	,0%	2,1%	1,9%	,0%	26,6%
Total	Count	602	27	44	56	46	25	800
	% within CNAE	75,3%	3,4%	5,5%	7,0%	5,8%	3,1%	100,0%
	% within Cluster Number of Case	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
	% of Total	75,3%	3,4%	5,5%	7,0%	5,8%	3,1%	100,0%

Fonte: Output do SPSS.

O objetivo é segmentar uma população de consumidores em grupos com características semelhantes sobre alguns aspectos do comportamento de pagamento da fatura mensal de energia elétrica. Portanto, buscou-se obter uma classificação de objetos baseada na experiência do pesquisador, retratada tanto pela seleção das variáveis, quanto e nas medidas de semelhança, que foram calculadas.

A identificação de padrões entre as variáveis de agrupamento estudadas traz a possibilidade de desenvolver perfis detalhados de grupos homogêneos de

consumidores de energia elétrica e o planejamento de táticas focadas nas características específicas de cada grupo de consumidores. Considerando os padrões de diferenças entre os agrupamentos e a clara distinção entre bons e maus pagadores evidenciados nas duas soluções, conclui-se que tanto a solução de dois, quanto de seis agrupamentos possuem um nível adequado de validade preditiva.

Em ambos os casos, obtiveram-se uma solução viável de agrupamento que representam conjuntos de consumidores com comportamentos de pagamento homogêneos. A solução de dois agrupamentos pode ser utilizada para fornecer um delineamento básico dos consumidores, nesse caso, sugerindo táticas gerais para consumidores com e sem dívida provisionadas à perdas (PCLD). Sendo assim, essa dívida de difícil recebimento deverá ser objeto de ações de cobrança específicas. A solução de seis agrupamentos pode ser empregada para táticas de cobrança mais complexas e focadas em segmentos com comportamentos e características altamente diferenciadas.

5.2.2 Análise de agrupamentos com base no histórico de pagamentos

Nessa segunda parte, avança-se na investigação do relacionamento entre consumidores e o histórico de pagamento verificado nos últimos seis meses. Busca-se então, agrupar os consumidores através das respostas (pagamentos) aos meios de cobrança existentes. Por exemplo, consumidores que usualmente pagam apenas após o corte, que pagam até a data de vencimento da fatura ou após receber uma carta de cobrança. Portanto, a associação de meios de pagamento a perfis de consumidores representa uma importante ferramenta para a concessionária definir estratégias com diferentes abordagens de cobrança e capaz de reduzir custos com operações inócuas àquele determinado segmento de consumidores.

Para uma população finita, composta pelos contratos de consumidores privados, a amostra utilizada de 1.774 contratos assegura um nível de confiança de 96% e uma margem máxima de erro desejada de 2,5%. A seleção da amostra foi obtida através da rotina *select samples* do SPSS. A amostra foi analisada com o objetivo de buscar dados atípicos, não sendo identificado casos que justifiquem uma eliminação de qualquer observação.

Nessa análise de agrupamento foram utilizadas seis variáveis métricas, relacionadas a seguir, que medem a quantidade de pagamentos realizados nos últimos 6 meses por modalidade de cobrança e representam a porcentagem de vezes em que cada consumidor pagou sua fatura respondendo àquele meio de cobrança específico.

V₁. Pagamento com o recebimento da fatura (qtd).

V₂. Pagamento com o recebimento do reaviso de cobrança (qtd).

V₃. Pagamento com a negativação (SPC/SERASA) (qtd).

V₄. Pagamento com visita de cobrança (qtd).

V₅. Pagamento com o corte do fornecimento (qtd)

V₆. Pagamento com assessoria de cobrança (qtd).

A medida de similaridade adotada para o conjunto de seis variáveis foi a distância euclidiana quadrada para obtenção da amplitude e do padrão de pagamentos realizados pelos consumidores. As variáveis estão na mesma escala, não sendo necessário normalizar os dados.

A representatividade da amostra e a multicolinearidade são analisadas visando assegurar uma adequada representação da estrutura dos dados.

O exame das seis variáveis usadas na análise de agrupamentos não apresentou distorções que justificassem a redução dos dados ou recálculo das medidas de distâncias. Assim, não foram encontradas evidências da existência de multicolinearidade que afetasse a análise.

5.2.2.1 Análise hierárquica de agrupamentos

O algoritmo de agrupamento utilizado foi o método de Ward. Nesse método, a similaridade entre dois agrupamentos é medida como a soma dos

quadrados entre os dois agrupamentos feita sobre todas as variáveis. De acordo com Hair e outros (2005), esse método tende a produzir agrupamentos de tamanhos mais homogêneos devido a sua minimização de variação interna.

Os resultados da análise de agrupamentos, incluindo os casos em combinação em cada estágio do processo e o coeficiente de aglomeração, são apresentados no ANEXO C. A quarta coluna representa o coeficiente de aglomeração obtido através da soma interna de quadrados entre os dois agrupamentos somados sobre todas as variáveis. O esquema de formação do dendrograma foi suprimido por ser extenso e por estar apresentado um esquema similar no ANEXO B.

Nessa etapa, avalia-se como os consumidores agrupados por segmento de cobrança respondem aos principais meios de cobrança utilizados pela Coelba. Portanto, os aspectos do comportamento de pagamento dos consumidores de energia elétrica aqui avaliados, aliados aos resultados obtidos na análise por atividade econômica, formam um conteúdo mais estruturado para subsidiar a definição das estratégias de relacionamento e cobrança. Do mesmo modo que no item 5.2.1, a análise hierárquica foi utilizada para obtenção de uma solução de quantos agrupamentos podem ser utilizados. Essa solução será validada através da análise não-hierárquica de agrupamentos.

Para determinar um número adequado de agrupamentos (regra de parada), foram utilizados os dez últimos estágios do processo de aglomeração apresentado no ANEXO C. Na Tabela 16, é apresentado um intervalo de soluções possíveis entre dois e dez agrupamentos, incluindo o coeficiente de aglomeração e a variação percentual no coeficiente de um para dois agrupamentos, e assim sucessivamente até o número de 10 agrupamentos.

O coeficiente de aglomeração sofre uma primeira grande alteração nas passagens de cinco para quatro (20,58%) e de quatro para três agrupamentos (22,34%). No entanto, as maiores variações ocorrem nas passagens de três para dois (41,76%) e de dois para um agrupamento (92,77%). Assim, as soluções de dois e três

agregados também são consideradas para avaliação da significância prática do agrupamento.

Tabela 16 - Análise do coeficiente de aglomeração para análise hierárquica de agrupamentos pelo método de Ward

Número de agrupamentos	Coeficiente de aglomeração	Varição Percentual no coeficiente para o próximo nível
10	93,05	8,43
9	100,89	9,07
8	110,05	10,06
7	121,12	12,03
6	135,69	13,22
5	153,62	20,58
4	185,23	22,34
3	226,62	41,76
2	321,26	92,77
1	619,31	-

Nota: Elaboração própria, adaptado do output SPSS.

5.2.2.2 Identificação de observações atípicas

Através do esquema de aglomeração pelo método de Ward (ANEXO C), obtido do pacote estatístico SPSS, é possível identificar observações atípicas que são representadas por um número pequeno de observações que se juntam muito tarde a um pequeno agrupamento.

No esquema de aglomeração apresentado no ANEXO C, as colunas ao lado direito do coeficiente apresentam o estágio em que o agrupamento surgiu pela primeira vez (*stage cluster first appears*) anotando os passos nos quais cada agrupamento foi formado. Uma observação que nunca foi reunida em um agrupamento tem o estágio zero. No estágio 5, ocorre a primeira junção com um outro agrupamento formado em um estágio anterior. Não foram localizadas observações individuais candidatas a dados atípicos, ou seja, que foram agrupadas tardiamente.

Os resultados obtidos através da regra de parada, ausência de dados atípicos e considerando perfis distintos de consumidores são adequados para uma solução de dois e de três agrupamentos. Portanto, a análise das duas soluções será levada adiante na análise não-hierárquica para “refinar” o estudo e auxiliar na decisão final.

5.2.2.3 Análise de agrupamentos não-hierárquica

Seguindo o mesmo roteiro apresentado no item 5.2.1, os procedimentos não-hierárquicos serão utilizados com o objetivo de validar os resultados obtidos na análise hierárquica.

Os dados foram processados no SPSS através do método *K-means*, utilizando dois e três agrupamentos sugeridos na análise hierárquica. Os valores dos centróides e o tamanho dos agrupamentos para as soluções de dois e três agrupamentos são apresentados na Tabela 17.

No método não-hierárquico, a solução de dois agrupamentos apresenta uma diferença significativa entre os agrupamentos. De fato, houve uma realocação dos consumidores que, mesmo apresentando a fatura como meio mais significativo de pagamento, também registra em menor escala, histórico de pagamento associados às demais modalidades de cobrança estudadas. Assim, tendo em vista a pouca correspondência e estabilidade das soluções apresentadas nos métodos hierárquico e não-hierárquico, não é possível confirmar a utilidade prática e teórica dos resultados.

Por outro lado, na solução de três agrupamentos também apresentada na Tabela 17, os perfis correspondem bem aos perfis do procedimento hierárquico. Apenas V₆ não apresenta diferença significativa entre os agrupamentos, o que explicado através da pequena proporção de consumidores sujeitos a essa modalidade de cobrança: assessoria de cobrança. Adicionalmente, os perfis de agrupamento aproximam-se bem, resultando em grupos com tamanhos parecidos.

Portanto, os resultados da solução de três agrupamentos apresentam correspondência e estabilidade entre os métodos hierárquico e não-hierárquico e são sujeitos à aceitação prática e teórica.

Tabela 17 - Solução de dois agrupamentos e de três agrupamentos da análise não-hierárquica com pontos sementes iniciais obtidos dos resultados hierárquicos

Agrupamento	Valores médios						Tamanho do agrupamento	
	V ₁	V ₂	V ₃	V ₄	V ₅	V ₆	K-means	Análise hierárquica
<i>Solução de dois agrupamentos - Centros do agrupamento final</i>								
1	0,8511	0,0545	0,0376	0,0273	0,0231	0,0066	789	1140
2	0,1108	0,0967	0,1413	0,1819	0,4227	0,0467	985	634
						total	1774	1774
<i>Solução de três agrupamentos - Centros do agrupamento final</i>								
1	0,9056	0,0344	0,0244	0,0152	0,0178	0,0027	683	639
2	0,2465	0,1731	0,2234	0,2170	0,1077	0,0324	520	501
3	0,0596	0,0432	0,0631	0,1358	0,6415	0,0569	571	634
						total	1774	1774

Nota: Elaboração própria, adaptado do output SPSS.

O teste de significância de diferenças entre centróides para as soluções com dois e três agrupamentos é apresentado na Tabela 18. Nas duas soluções de agrupamentos, todas as variáveis apresentaram diferença significativa entre os agrupamentos.

Tabela 18 - Teste de significância de diferenças entre centros de agrupamento para dois e três agrupamentos

Solução de dois agrupamentos							
	Cluster		Error		F	Sig.	
	Mean Square	df	Mean Square	df			
FATURA	240,040	1	,026	1772	9064,017	,000	
REAVISO	,777	1	,018	1772	42,107	,000	
NEGATIV	4,711	1	,023	1772	205,106	,000	
VISITA	10,479	1	,030	1772	354,159	,000	
CORTE	69,958	1	,058	1772	1212,534	,000	
ASSESSO	,707	1	,010	1772	71,048	,000	

Solução de três agrupamentos							
	Cluster		Error		F	Sig.	
	Mean Square	df	Mean Square	df			
FATURA	125,071	2	,021	1771	6014,842	,000	
REAVISO	3,345	2	,015	1771	221,070	,000	
NEGATIV	6,280	2	,019	1771	338,599	,000	
VISITA	6,232	2	,028	1771	218,775	,000	
CORTE	67,413	2	,021	1771	3194,857	,000	
ASSESSO	,462	2	,010	1771	47,018	,000	

Fonte: Output do SPSS.

Ao avaliar na amostra o desempenho dos principais meios de cobrança utilizados na concessionária, considerou-se como meio de cobrança predominante aquele associado a mais da metade dos últimos seis pagamentos efetuados. A avaliação considerou os segmentos de cobrança propostos no estudo de segmentação apresentado no item 5.1 deste trabalho. Análise aqui efetuada foi baseada na Tabela 19 e considerou um fluxo de cobrança típico adotado na companhia.

Tabela 19 - Comportamento de pagamento por tipo de cobrança e segmento de cobrança

Segmento de Cobrança		Modalidade de Cobrança					
		V1 FATURA	V2 REAVISO	V3 NEGATIVAÇÃO	V4 VISITA	V5 CORTE	V6 ASS. COBRANÇA
Privados Consumo Baixo	PCB	47,65%	7,18%	8,23%	8,54%	22,55%	5,85%
Privados Consumo Pequeno	PCP	41,28%	8,46%	10,65%	11,83%	25,63%	2,14%
Privados Consumo Médio	PCM	42,84%	7,50%	8,29%	15,36%	25,04%	0,96%
Privados Consumo Alto	PCA	51,03%	5,86%	8,05%	11,95%	22,87%	0,23%
Coelba		44,01%	7,79%	9,52%	11,31%	24,49%	2,88%

Notas: foi considerado os últimos 6 pagamentos realizados.

Elaboração própria.

Ao analisar a Tabela 19, verifica-se que a “fatura” é a modalidade mais efetiva para 44,01% dos consumidores, destacando-se o segmento PCA (alto consumo), com 51,03% dos consumidores respondendo a esse meio de cobrança. Para os segmentos PCP (consumidores de pequeno porte) e PCM (consumidores de médio porte) a fatura também é a principal modalidade de cobrança, sendo mais efetiva para 41,28% e 42,84%, respectivamente. Ressalta-se que a “fatura”, diferentemente das demais modalidades de cobrança, é entregue indistintamente a todos os consumidores a cada mês. As demais modalidades analisadas a seguir são seletivas e dependem da política de cobrança em vigor e das restrições impostas pelo Agente Regulador.

A segunda modalidade de cobrança mais efetiva é o “corte do fornecimento”. Com 24,49% do total de pagamentos associados a essa modalidade, destacam-se os consumidores dos segmentos PCP e PCM com valores variando entre 25,63% e 25,04%. O segmento PCP apresenta um desempenho menor, de 22,55%.

Apesar de considerar o efeito da cobrança através da negativação medido até o início da próxima modalidade de cobrança, neste caso, a visita, é importante considerar o seu caráter suplementar tendo em vista que a maior probabilidade de pagamento está associada ao momento de aquisição de crédito em um estabelecimento comercial ou bancário.

A última modalidade de cobrança, a “assessoria de cobrança” preenche uma lacuna importante na recuperação de créditos que normalmente já foram

submetidos a um ou mais tipos de cobrança, porém sem efetividade. Logo, é bastante seletivo e contempla um universo bastante restrito de consumidores em comparação com as demais modalidades. Mesmo assim, 2,88% dos consumidores efetuaram pelo menos um pagamento, entre os seis últimos, motivados por esse tipo de cobrança. Destaca-se o segmento de baixo consumo (PCP), que apresenta um desempenho de 5,85%.

A avaliação dos segmentos de cobrança e do comportamento de pagamento associado, objetiva aprofundar a análise do comportamento dos consumidores e obter agregados de consumidores com perfis semelhantes quanto aos meios de cobrança que motivaram o pagamento da fatura mensal de energia elétrica. A seleção das variáveis e as medidas de similaridade obtidas são as peças-chave na construção do modelo de segmentação por comportamento de pagamento.

Com base nos dados da Tabela 17, que apresentam a solução de dois e três agrupamentos da análise não-hierárquica, são apresentados a seguir os perfis de agrupamentos sugeridos pelo modelo.

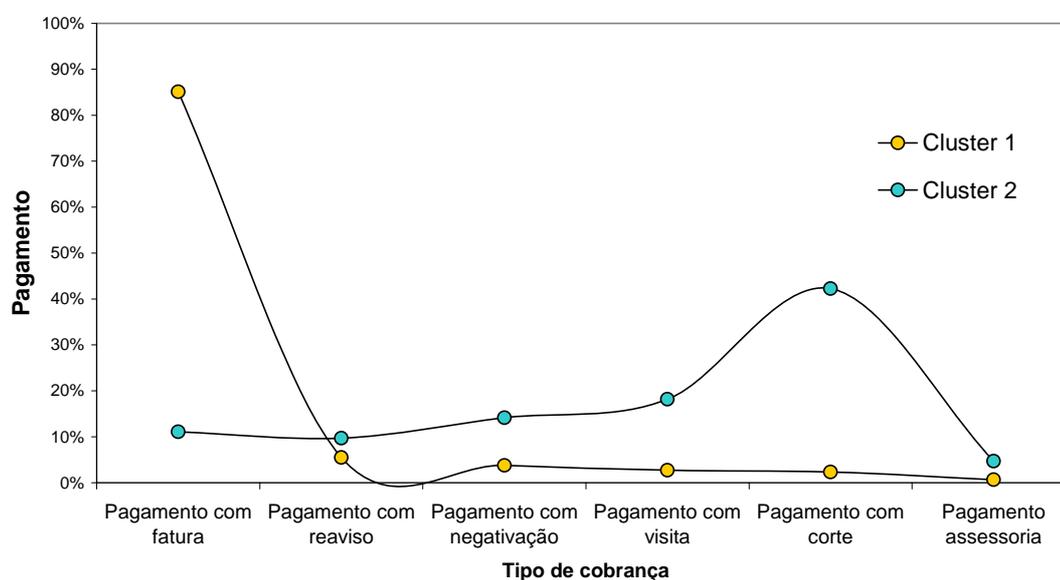


Figura 18 - Perfis de agrupamento para solução de dois agregados
Nota: Elaboração própria.

Na Figura 18 foi formado o perfil para os dois agrupamentos considerados significantes em termos práticos e teóricos. Verifica-se que o *cluster 1* é significativamente maior quanto à dimensão da variável V_1 (fatura), ao passo que o

cluster 2 possui maiores valores associados à variável V_5 (corte). No *cluster 1* foram agrupados os consumidores com maior percepção da “fatura” como o principal modalidade de cobrança, enquanto no *cluster 2*, estão concentrados aqueles com maior percepção do “corte” como principal modalidade de cobrança.

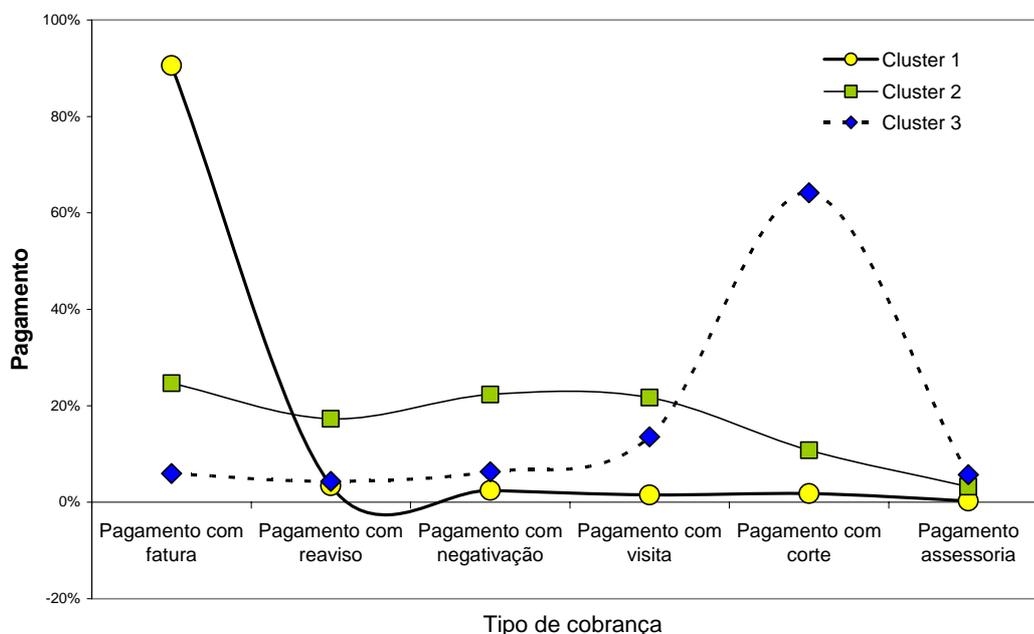


Figura 19 - Perfis de agrupamento para solução de três agregados
Nota: Elaboração própria.

O segundo perfil é apresentado na Figura 19. Foi formado o perfil para os três agrupamentos considerados significantes em termos práticos e teóricos. Verifica-se que o *cluster 1* é significativamente maior quanto à dimensão da variável V_1 (fatura), ao passo que o *cluster 3* possui maiores valores associados à variável V_5 (corte). No *cluster 1* foram agrupados os consumidores com maior percepção da “fatura” como o principal modalidade de cobrança, enquanto no *cluster 3*, estão concentrados aqueles com maior percepção do “corte” como principal modalidade de cobrança. No *cluster 2* foram agrupados os consumidores com uma percepção de distribuída de uma forma mais proporcional, indicando um comportamento de pagamento muito irregular. Apenas a variável V_6 (assessoria) não é significativamente diferente entre os três agregados, explicado por causa da própria modalidade de cobrança, que consiste na recuperação de créditos, uma vez esgotadas as demais ações de cobrança. Logo, um volume proporcionalmente menor de consumidores paga suas faturas em função da assessoria de cobrança.

A análise de agrupamentos possibilitou uma segmentação dos consumidores com base no comportamento de pagamento, apontando as percepções dos segmentos frente aos estímulos utilizados pela concessionária para acelerar o pagamento da fatura mensal de energia elétrica.

A solução de três agrupamentos fornece segmentos com tamanhos substanciais e sem observações atípicas. Por representar conjuntos de consumidores com percepções homogêneas sobre as modalidades de cobrança, é possível identifica-los submetê-los a estratégias e operações de cobrança adequadas ao perfil do consumidor.

5.3 MEDIÇÃO DO RISCO DE INADIMPLÊNCIA PARA A CARTEIRA DE CONSUMIDORES DA COELBA

No item anterior, buscou-se responder ao primeiro problemas da pesquisa ao avaliar opções de agrupamentos com base nos aspectos da dívida dos consumidores e nos histórico de pagamentos. Os resultados encontrados formam um insumo para a próxima etapa da pesquisa.

O segundo e último problema da pesquisa objetiva responder qual a medida numérica da incerteza relacionada ao recebimento de uma carteira de valores a receber de consumidores de energia elétrica (risco de inadimplência). Com esse objetivo, será utilizado um modelo de regressão logística binária para prever a existência ou não de inadimplentes.

Segundo Hair e outros (2005, p.231), a aplicação da regressão logística é mais adequada quando a variável dependente é binária, além de não depender das suposições rígidas de normalidade multivariada e de igualdade de matrizes de variância-covariância, exigidas na análise discriminante. A construção do modelo logístico buscará prever a presença ou ausência de inadimplência com base em um conjunto de variáveis independentes que melhor caracterizem o perfil dos consumidores da Coelba. Os coeficientes logísticos serão utilizados para estimar as proporções da probabilidade para cada uma das variáveis independentes no modelo.

Usando procedimento logístico binário da regressão, objetiva-se criar um modelo de *score* comportamental (*behavioural scoring*) identificando-se os consumidores com maior propensão a inadimplir e priorizar as ações de cobrança adequadas. Assim, a empresa concessionária poderá alterar a política de cobrança reduzindo as perdas com devedores incobráveis e os custos operacionais. Trata-se de um primeiro passo na direção de otimizar as operações de cobrança.

A pesquisa consiste em três etapas-chave: selecionar as variáveis dependente e independentes, avaliar a adequação do tamanho da amostra para análise planejada e dividir a amostra para fins de validação.

5.3.1 Seleção de variáveis dependente e independentes

Os dados selecionados para condução deste trabalho, que objetiva classificar consumidores inadimplentes e adimplentes, utilizam como base uma amostra da carteira de consumidores da concessionária.

Foi analisado o histórico de pagamentos das faturas dos últimos seis meses. A informação do sexto mês foi utilizada para definir a situação do cliente, se adimplente ou inadimplente. As informações de pagamentos dos cinco meses restantes são utilizadas para testar a capacidade preditiva do modelo para acertar a situação do cliente verificada no sexto mês.

Baseado nos dados de pagamento destes clientes foi definida a variável dependente, Situação do Cliente "INAD", que assume valores 0 (adimplente) e 1 (inadimplente), de acordo com a seguinte classificação:

- a) Cliente adimplente - clientes com até um pagamento efetuado com até 10 dias de atraso, nos últimos cinco meses.
- b) Cliente inadimplente - cliente com mais de um pagamento com atraso superior a 10 dias, nos últimos cinco meses.

O conjunto inicial de variáveis selecionadas é mostrado a seguir. Foram obtidas a partir das informações disponíveis na empresa concessionária.

- 1) Consumo médio mensal (kWh). "KWH_MED"
- 2) Fraude de energia nos últimos seis meses (sim/não). "FAT_FRA"
- 3) Pagamento 1: Quantidade de dias após o vencimento. "F1"
- 4) Pagamento 2: Quantidade de dias após o vencimento. "F2"
- 5) Pagamento 3: Quantidade de dias após o vencimento. "F3"
- 6) Pagamento 4: Quantidade de dias após o vencimento. "F4"
- 7) Pagamento 5: Quantidade de dias após o vencimento. "F5"

A amostra utilizada é a mesma utilizada na análise de agrupamentos no item 5.2.2. Foi utilizada uma amostra de 1.774 consumidores do setor privado para realização da análise de regressão logística, sendo 1.109 inadimplentes e 665 adimplentes. Essa amostra assegura um nível de confiança de 96% e uma margem máxima de erro desejada de 2,5%. A seleção da amostra foi obtida através da rotina *select samples* do SPSS.

5.3.2 Análise dos dados

O conjunto de consumidores para validação foi gerado aleatoriamente no pacote estatístico SPSS Versão 13.0, utilizando-se a distribuição de Bernoulli como um parâmetro 0,7. A nova variável atribui valores 0 ou 1 de acordo com a probabilidade especificada. Como resultado, 1.208 consumidores, equivalente a 68,1% do total, ficaram com o valor de validação de 1. Esses 1.208 consumidores serão usados para criar o modelo de regressão logística. Os 566 consumidores restantes ficaram com um valor de validação 0 e passam a compor uma sub-amostra que será empregada para validar os resultados modelo.

Foi estruturada uma base de dados agregando os valores numéricos referentes ao conjunto das possíveis variáveis explicativas pré-selecionadas para utilização na construção dos modelos. Com esse modelo, foram estabelecidas

correlações parciais para cada variável e a variável mais discriminante pode ser escolhida de acordo com o critério de seleção (HAIR e outros, 2005).

Para os dados do estudo, foram gerados resultados utilizando o método *stepwise forward:LR*. Nesse método, as variáveis independentes entram sequencialmente no modelo de acordo com o poder discriminatório que elas acrescentam à previsão de pertinência no grupo.

Após ter construído o modelo, é necessário determinar se ele aproxima-se razoavelmente do comportamento dos dados em estudo. De acordo com Hair e outros (2005, p.234), para testar o ajuste do modelo em termos de pertinência a grupos, o procedimento de regressão logística binária utiliza o método de matrizes de classificação e o teste de *Hosmer-Lemeshow*.

No teste de *Hosmer-Lemeshow*, os casos são primeiramente divididos em aproximadamente dez classes iguais. Em seguida, os números de eventos reais e previstos são comparados em cada classe com a estatística qui-quadrado. Com esse teste, obtém-se uma medida de precisão preditiva com base na real previsão da variável dependente, ao invés de utilizar o valor de verossimilhança.

As matrizes de classificação consistem em tabelas que comparam a classificação realizada pelos modelos desenvolvidos com a classificação original das observações da amostra. Essas matrizes de classificação são construídas através da análise de cada observação com o objetivo de conhecer se ela foi corretamente classificada pelos modelos. Os resultados dessa análise são os percentuais de acerto e erro de classificação dos modelos.

Segundo Hair (p. 264, 2005), a medida final de ajuste do modelo é o valor de *Hosmer and Lemeshow Test*, o qual mede a correspondência entre valores reais e os previstos da variável dependente.

A medida *Hosmer-Lemeshow* de ajuste geral tem um teste estatístico que indica que não houve diferença estatisticamente significativa entre as classificações observadas e previstas para todos os modelos com duas ou mais variáveis.

O valor Hosmer-Lemeshow mede a correspondência dos valores efetivos e previstos da variável dependente. Neste caso, o melhor ajuste do modelo é indicado por uma diferença menor na classificação observada e prevista. Um bom ajuste de modelo é indicado por um valor chi-quadrado não significativo (HAIR e outros, 2005). O teste de *Hosmer-Lemeshow* é mostrado na Tabela 20, a seguir. Os graus de liberdade (DF) determinam a possibilidade de combinações ao acaso dos dados.

Tabela 20 - Teste de Hosmer-Lemeshow

Step	Chi-square	df	Sig.
1	53,799	6	,000
2	57,653	7	,000
3	168,678	8	,000
4	154,316	8	,000
5	160,669	8	,000

Fonte: Output do SPSS.

Essas medidas sugerem a aceitação do modelo do último passo como um modelo significativo de regressão logística. Esta estatística é o teste mais seguro de ajuste modelo para regressão logística binária, porque se agrega as observações em grupos de casos "semelhantes." A estatística é calculada baseada então nestes grupos.

Após o teste de *Hosmer and Lemeshow*, é utilizado o Método de Matrizes de Classificação para avaliar a precisão preditiva do modelo de regressão logística.

Na Tabela 21, a Matriz de Classificação mostra a classificação dos casos preditos pelo modelo comparando com os dados reais referentes à variável "INAD" (inadimplente).

Observa-se que o modelo apresenta uma explicação ligeiramente melhor para os clientes com a situação de inadimplente, com o percentual de acerto de 84,8% na amostra utilizada para gerar o modelo e 86,1% na amostra de validação.

De uma forma geral, modelo classifica corretamente 84,0% dos clientes em um dos dois grupos. Portanto, apresentando um expressivo percentual representativo de convergência. Para cada caso, a resposta predita é SIM se a

probabilidade predita de inadimplência for maior que o valor de corte de 0,5 especificado no modelo.

Tabela 21 - Tabela de classificação com os resultados práticos do modelo de regressão logística

		Classification Table ^d						
		Predicted						
		Selected Cases ^a				Unselected Cases ^{b,c}		
		INAD		Percentage Correct	INAD		Percentage Correct	
NÃO	SIM	NÃO	SIM					
Observed								
Step 1	INAD	NÃO	386	78	83,2	166	59	73,8
		SIM	139	493	78,0	73	222	75,3
	Overall Percentage				80,2			74,6
Step 2	INAD	NÃO	384	80	82,8	172	53	76,4
		SIM	107	525	83,1	46	249	84,4
	Overall Percentage				82,9			81,0
Step 3	INAD	NÃO	387	77	83,4	170	55	75,6
		SIM	100	532	84,2	44	251	85,1
	Overall Percentage				83,9			81,0
Step 4	INAD	NÃO	388	76	83,6	170	55	75,6
		SIM	98	534	84,5	43	252	85,4
	Overall Percentage				84,1			81,2
Step 5	INAD	NÃO	385	79	83,0	171	54	76,0
		SIM	96	536	84,8	41	254	86,1
	Overall Percentage				84,0			81,7

a. Selected cases 'VALIDATE EQ 1

b. Unselected cases 'VALIDATE NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,500

Fonte: Output do SPSS.

A amostra utilizada para validar os resultados do modelo propicia uma verificação adicional da capacidade preditiva do modelo. Estes resultados são mostrados na coluna *Unselected Cases* da Tabela 21. Na amostra de validação, ao todo, 81,7% dos consumidores foram classificados corretamente pelo modelo. Esse resultado sugere que, ao todo, o modelo está de fato correto.

Ao observar na Tabela 21 os dados referentes aos casos não selecionados para a criação do modelo e utilizados como amostra de validação, verifica-se que 76,0% dos casos foram enquadrados corretamente para a situação de cliente adimplente e 86,1% para a situação de cliente inadimplente, totalizando uma média de 81,7% de acerto geral.

Na Tabela 22, a seguir, são apresentadas as variáveis presentes na equação gerada pela ferramenta de regressão logística, bem como os seus betas e a significância das variáveis independentes.

Após diversas interações, o modelo final selecionou 5 das 7 variáveis incluídas inicialmente no modelo. Excluiu as variáveis: “Consumo médio mensal (kWh)” e “Fraude de energia nos últimos seis meses (sim/não)”. As variáveis que resultaram do modelo final são: “Pagamento 1”, “Pagamento 2”, “Pagamento 3”, “Pagamento 4” e “Pagamento 5”.

Na Tabela 22, são sumarizados os efeitos de cada variável. Os coeficientes estimados para as variáveis independentes e a constante foram avaliados através da estatística Wald em termos de significância estatística. Se os níveis de significância da estatística de Wald forem menores que 0,05, então a variável é incorporada ao modelo. Nesse estudo, os coeficientes são estatisticamente significantes no nível de 0,01, embora a constante não seja significativa. Assim, as variáveis individuais são significantes e foram interpretadas individualmente.

Tabela 22 - Detalhamento das medidas de exame das variáveis incluídas no modelo de regressão logística

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	F1	,104	,007	234,656	1	,000	1,110
	Constant	-1,071	,100	114,625	1	,000	,343
Step 2	F1	,072	,007	98,647	1	,000	1,074
	F5	,056	,006	86,304	1	,000	1,058
	Constant	-1,658	,125	176,876	1	,000	,190
Step 3	F1	,056	,008	51,757	1	,000	1,058
	F2	,029	,006	20,995	1	,000	1,029
	F5	,048	,006	58,953	1	,000	1,049
	Constant	-1,792	,131	187,180	1	,000	,167
Step 4	F1	,055	,008	47,500	1	,000	1,056
	F2	,023	,007	11,857	1	,001	1,023
	F3	,012	,005	7,273	1	,007	1,013
	F5	,045	,006	50,743	1	,000	1,046
	Constant	-1,844	,133	190,807	1	,000	,158
Step 5	F1	,054	,008	45,136	1	,000	1,055
	F2	,016	,007	5,280	1	,022	1,016
	F3	,010	,004	4,931	1	,026	1,010
	F4	,014	,006	5,705	1	,017	1,014
	F5	,042	,007	39,672	1	,000	1,043
	Constant	-1,873	,135	192,694	1	,000	,154

a. Variable(s) entered on step 1: F1.

b. Variable(s) entered on step 2: F5.

c. Variable(s) entered on step 3: F2.

d. Variable(s) entered on step 4: F3.

e. Variable(s) entered on step 5: F4.

Fonte: Output do SPSS.

As variáveis preditoras e os coeficientes são mostrados no último passo (*step* 5) da Tabela 22 foram utilizadas para fazer as previsões de inadimplência.

Os coeficientes “B” são medidas de variação na proporção das probabilidades, “Exp (B)” é a sua transformação logarítmica. Se “B” é positivo, então sua transformação (anti-logaritmo) é maior que 1, significando que aumenta a probabilidade de inadimplência. Por outro lado, quando “B” é negativo, a sua transformação é menor que 1, o que diminui a probabilidade prevista (HAIR e outros, 2005). Assim, “Exp (B)” representa taxa de mudança nas probabilidades da inadimplência quando ocorre uma mudança de uma unidade na variável preditora. Nesse caso, em termos de probabilidade, essa diferença é significativa porque a probabilidade original para um consumidor que usualmente paga suas faturas sem a necessidade de outras modalidades de cobrança é próxima de zero.

Uma forma de medir a qualidade do ajuste do modelo é estatística chamada distância de Cook. Essa estatística é uma medida sintética da influência de um único caso baseada nas mudanças totais em todos os outros resíduos quando o caso é eliminado do processo de estimação. Valores elevados, geralmente maiores que um, indicam que o caso exerce uma substancial influência afetando os coeficientes de regressão estimados. Na Figura 20, são mostradas as distâncias de Cook para os casos utilizados nesse estudo. A forma da distância de Cook geralmente concentra uma elevada quantidade de consumidores na base, com algumas exceções que ficam dispersas. Essas exceções são pontos com alta alavancagem e podem influenciar a análise. Os dez casos verificados com uma distância de Cook igual ou maior que 1 são clientes que tornaram-se adimplentes, embora existam registros de pagamentos efetuados com mais de 180 dias de atraso. No entanto, essa pequena concentração não foi considerada como valores capazes de influenciar substancialmente a análise.

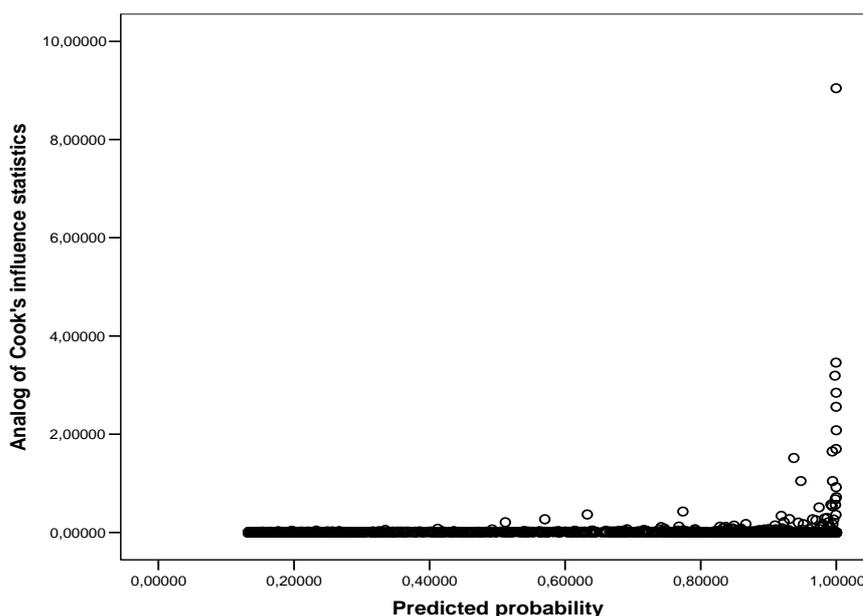


Figura 20 - Distância de Cook em relação à probabilidade predita no modelo logístico

Fonte: Output do SPSS.

No SPSS 13.0, foi aplicado o modelo de regressão logística binário, método *stepwise forward:LR*, para definir o modelo final que minimiza o número de variáveis e maximiza a precisão do modelo. Inicialmente, esse modelo não inclui nenhuma variável preditora, conforme mostrado na Tabela 23. Em cada passo, a

variável preditora com maior escore e cujo valor da significância é menor que 0,05, é adicionada ao modelo.

Tabela 23 - Variáveis predictoras na partida da regressão logística

			Score	df	Sig.
Step	Variables	F1	257,610	1	,000
0		F2	224,795	1	,000
		F3	189,938	1	,000
		F4	160,891	1	,000
		F5	206,021	1	,000
		KWH_MED	,094	1	,760
		FAT_FRA	,077	1	,781
	Overall Statistics		338,209	7	,000

Fonte: Output do SPSS.

Todas as variáveis que ficaram fora da análise até o último passo possuem um valor de significância maior que 0,05, logo não são mais adicionadas ao modelo. Essas variáveis são mostradas na Tabela 24.

Tabela 24 - Variáveis preditoras que ficaram fora do modelo de regressão logística

Variables not in the Equation					
			Score	df	Sig.
Step 1	Variables	F2	57,341	1	,000
		F3	59,115	1	,000
		F4	54,867	1	,000
		F5	105,099	1	,000
		KWH_MED	,375	1	,540
		FAT_FRA	,018	1	,894
	Overall Statistics		131,839	6	,000
Step 2	Variables	F2	20,126	1	,000
		F3	19,489	1	,000
		F4	13,335	1	,000
		KWH_MED	,051	1	,821
		FAT_FRA	,006	1	,936
			Overall Statistics		30,109
Step 3	Variables	F3	8,538	1	,003
		F4	5,590	1	,018
		KWH_MED	,142	1	,706
		FAT_FRA	,012	1	,912
			Overall Statistics		12,211
Step 4	Variables	F4	4,039	1	,044
		KWH_MED	,178	1	,674
		FAT_FRA	,065	1	,799
			Overall Statistics		4,230
Step 5	Variables	KWH_MED	,115	1	,734
		FAT_FRA	,090	1	,764
			Overall Statistics		,203

Fonte: Output do SPSS.

As variáveis escolhidas pelo modelo tiveram mudanças significativas na estatística $-2 \log (-2LL)$, que é a medida geral de quão bem o modelo está ajustado. As mudanças no $-2LL$ representam a variação no ajuste preditivo de uma equação para outra. No modelo avaliado, as medidas de ajuste melhoraram com a inclusão das variáveis. Na Tabela 25 é mostrado o histórico de interação da regressão. Valores menores da medida $-2LL$ indicam um melhor ajuste do modelo. A cada passo, quando uma nova variável é incluída no modelo, a estatística de probabilidade $-2 \log$ diminui indicando uma melhora no modelo. Nesse modelo, foram incluídas três variáveis, sendo que o valor $-2LL$ foi reduzido a partir do valor base de 1.493,525 para 869,119.

Tabela 25 – Histórico de interação do modelo de regressão logística

Iteration History^{a,b,c,d}

Iteration		-2 Log likelihood	Coefficients					
			Constant	F1	F5	F2	F3	F4
Step 1	1	1174,776	-,444	,040				
	2	1062,508	-,873	,077				
	3	1044,069	-1,038	,099				
	4	1043,304	-1,069	,104				
	5	1043,302	-1,071	,104				
	6	1043,302	-1,071	,104				
Step 2	1	1103,038	-,602	,029	,015			
	2	945,525	-1,231	,054	,034			
	3	911,254	-1,553	,068	,049			
	4	908,737	-1,650	,071	,055			
	5	908,719	-1,658	,072	,056			
	6	908,719	-1,658	,072	,056			
Step 3	1	1083,821	-,667	,022	,012	,012		
	2	923,047	-1,321	,042	,028	,020		
	3	887,332	-1,673	,053	,042	,027		
	4	884,652	-1,782	,056	,047	,029		
	5	884,632	-1,792	,056	,048	,029		
	6	884,632	-1,792	,056	,048	,029		
Step 4	1	1078,402	-,686	,021	,010	,009	,005	
	2	915,254	-1,353	,041	,026	,016	,009	
	3	878,694	-1,719	,051	,039	,021	,012	
	4	875,935	-1,834	,054	,044	,022	,012	
	5	875,914	-1,844	,055	,045	,023	,012	
	6	875,914	-1,844	,055	,045	,023	,012	
Step 5	1	1071,254	-,709	,021	,009	,008	,004	,005
	2	909,253	-1,370	,040	,023	,012	,007	,009
	3	872,056	-1,742	,050	,036	,015	,009	,012
	4	869,143	-1,862	,053	,041	,016	,010	,014
	5	869,119	-1,873	,054	,042	,016	,010	,014
	6	869,119	-1,873	,054	,042	,016	,010	,014

Fonte: Output do SPSS.

As medidas *Cox & Snell*, *Nagelkerke* e “pseudo” R^2 também têm o intuito de apresentar o ajuste geral do modelo. Segundo Hair e outros (2005, p.264), estas medidas de adequação de ajuste comparam as probabilidades estimadas com as probabilidades observadas, sendo que valores mais altos significam um melhor ajuste do modelo.

A medida R^2 Cox & Snell está limitada, no sentido em que não pode alcançar o valor máximo de 1; e, conseqüentemente, *Nagelkerke* propôs uma modificação que apresenta o domínio de 0 a 1.

A estatística pseudo R^2 é baseada na comparação da medida de similaridade do modelo atual com o modelo “nulo”. A seguir são apresentadas três medidas comparáveis com a medida R^2 utilizada em regressão múltipla.

Os resultados do modelo gerado são mostrados na Tabela 26. O valor de Cox e Snell é 0,434, o valor de Nagelkerke é 0,584. Observa-se que conforme foram incluindo as variáveis, o modelo foi melhor se ajustando, de forma a apresentar valores mais altos.

Tabela 26 - Resultados das medidas de R^2 de Cox e Snell, de Nagelkerke e “pseudo” R^2

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1043,302 ^a	,337	,453
2	908,719 ^a	,413	,556
3	884,632 ^a	,426	,573
4	875,914 ^a	,431	,579
5	869,119 ^a	,434	,584

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Fonte: Output do SPSS.

O valor do “pseudo” R^2 é 0,418, calculado através da equação 4.9 com dados da Tabela 25, do seguinte modo:

$$R^2_{\text{logit}} = \frac{-2LL_{\text{nulo}} - (-2LL_{\text{modelo}})}{-2LL_{\text{nulo}}} \quad (4.9)$$

$$= \frac{1.493,525 - 869,119}{1.493,525} = 0,418$$

O valor do “pseudo” R^2 varia de um mínimo de 0 para um máximo de 1. Quanto maior for o “pseudo” R^2 , mais da variação é explicada pelo modelo.

Com o objetivo de fazer mais uma verificação adicional, foi construído um modelo utilizando o método *stepwise* inverso. O método inverso começa com um modelo que inclui inicialmente todas as variáveis preditoras. Em cada etapa, a variável que contribui menos é removida do modelo, até que todas as variáveis no modelo sejam significativas. Observa-se que os dois métodos escolheram as mesmas variáveis. Assim, pode-se afirmar que o modelo é confiável. A Tabela 27 apresenta um resumo dos resultados do método *stepwise backward:LR*, no modelo de regressão logística. Esse método é o inverso do *stepwise forward:LR*, porque considera inicialmente todas as variáveis no modelo.

Tabela 27 - Resultados do método *stepwise* inverso no modelo de regressão logística

Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change	
Step 1	F1	-461,477	54,037	1	,000
	F2	-437,335	5,754	1	,016
	F3	-437,291	5,665	1	,017
	F4	-437,837	6,757	1	,009
	F5	-457,366	45,816	1	,000
	KWH_MED	-434,515	,113	1	,737
	FAT_FRA	-434,502	,088	1	,767
Step 2	F1	-461,568	54,132	1	,000
	F2	-437,368	5,731	1	,017
	F3	-437,310	5,615	1	,018
	F4	-437,868	6,732	1	,009
	F5	-457,567	46,129	1	,000
	KWH_MED	-434,560	,115	1	,734
Step 3	F1	-461,589	54,059	1	,000
	F2	-437,393	5,667	1	,017
	F3	-437,352	5,585	1	,018
	F4	-437,957	6,794	1	,009
	F5	-457,801	46,483	1	,000

Fonte: Output do SPSS.

Portanto, essas medidas combinadas sugerem a aceitação do modelo do último passo como um modelo significativo de regressão logística.

Esses resultados representam uma expressiva adequação do modelo para explicar o comportamento da variável dependente “Situação do Cliente”.

A partir do resultado de regressão logística, obtêm-se a seguinte equação final:

$$\text{SIT_CLIENTE (INAD)} = - 1,873 + (- 0,054 \times \text{F1}) + (0,016 \times \text{F2}) + (0,010 \times \text{F3}) + (0,014 \times \text{F3}) + (0,042 \times \text{F5})$$

Onde: (5.1)

SIT_CLIENTE = Situação do Cliente (ou inadimplente “1”, ou adimplente “0”). Ponto de corte = 0,5. Logo, se $\geq 0,5$, então = inadimplente.

F1 = quantidade de dias entre o vencimento e o pagamento da fatura 1.

F2 = quantidade de dias entre o vencimento e o pagamento da fatura 2.

F3 = quantidade de dias entre o vencimento e o pagamento da fatura 3.

F4 = quantidade de dias entre o vencimento e o pagamento da fatura 4.

F5 = quantidade de dias entre o vencimento e o pagamento da fatura 5.

Desta forma, a função logística que melhor discrimina os consumidores inadimplentes do setor privado é dada pela equação 4.7.

$$f(z) = \frac{1}{1 + e^{-z}} \quad (4.7)$$

Onde,

$e \approx 2,718$;

O modelo de risco de crédito desenvolvido neste trabalho forneceu uma precisão preditiva satisfatória na classificação de clientes inadimplentes ou do risco de inadimplência.

Esse modelo selecionou duas variáveis significantes no nível 0,01, de um total de seis. Esse modelo funciona ligeiramente melhor para prever a inadimplência do que a adimplência, sendo 84,8% contra 83,0%. Para um modelo cujo objetivo principal é prever a inadimplência esse é um dado favorável. Ao todo, o modelo prevê precisamente a inadimplência ou adimplência de 84,0% das vezes na amostra de análise e 81,7% na amostra de validação, o que um resultado muito bom para modelos de *Credit Scoring*.

6 CONCLUSÃO

6.1 CONSIDERAÇÕES FINAIS

Caouette, Altman e Naraynan (2000), ressaltam que sempre há um elemento de risco quando uma pessoa adquire um produto ou serviço sem pagar imediatamente por ele. Especificamente, uma concessionária de distribuição de energia elétrica aceita o risco de inadimplência de todos os seus consumidores. De acordo com esses autores, o crédito é uma expectativa de receber uma quantia em dinheiro, dentro de um espaço de tempo limitado. Logo, o risco de crédito, mais especificamente o risco de inadimplência, é a chance de que essa expectativa não se realize, descontadas as expectativas de recuperação com outras medidas de cobrança.

Indo mais adiante, Crouhy, Galai e Mark (2004) enfatizam que a principal razão para uma instituição não-financeira adotar modelos de gerência de risco é a identificação de fatores de risco que afetam a volatilidade de suas receitas. Portanto, devem considerar os efeitos desse risco de crédito sobre o planejamento e o orçamento, com reflexos diretos na conta de lucros e perdas.

Segundo o atual contexto da regulação do serviço público de distribuição de energia elétrica, foram analisados os principais aspectos associados com a inadimplência e os custos operacionais e financeiros dela decorrentes.

Ressalta-se que na Nota Técnica nº. 349/ 2007 - SRE / ANEEL, foi ratificado que somente devem fazer parte das tarifas as receitas irrecuperáveis, definidas conforme critérios que levaram em consideração aspectos sócio-econômicos e demográficos da área de concessão. Assim, os custos com receitas irrecuperáveis acima do limite fixado, não são reconhecidos pela ANEEL. Permanece a determinação de que os custos decorrentes das operações de cobrança devem ser arcados pelo consumidor inadimplente. No entanto, os valores autorizados para a cobrança da taxa de religação são inferiores aos custos praticados. Portanto, o quadro institucional dar sinais contundentes às concessionárias de distribuição da necessidade da contínua melhoria da eficiência produtiva.

O objetivo deste trabalho foi o de desenvolver um modelo preditivo de inadimplência para consumidores de energia elétrica, através da análise do comportamento de pagamento dos consumidores de energia elétrica da Coelba e cálculo de equações de regressão logística que explicam a probabilidade da inadimplência.

Os desafios da pesquisa eram de responder as duas questões: i) quais os meios de cobrança mais adequados ao perfil dos agrupamentos de consumidores inadimplentes? e ii) qual a medida numérica da incerteza relacionada ao recebimento de uma carteira de valores a receber de consumidores de energia elétrica (risco de inadimplência)? As duas questões foram respondidas através de modelos estatísticos utilizando análise agrupamentos e análise de regressão logística.

Atuando em um ambiente de restrições de caráter financeiro, operacional e regulatório, a metodologia proposta mostrou-se viável e apropriada para avaliação em massa dos consumidores da Coelba, possibilitando gerar as carteiras diárias de cobrança de acordo com as diferenças entre cada segmento de consumidores e com a variância interna própria existente em cada bloco de consumidores.

Um estudo de comportamento de pagamento, utilizando a técnica estatística de análise de agrupamentos, foi desenvolvido para associar os atuais meios de cobrança aos segmentos e consumidores segundo a efetividade de cada ação de cobrança, quais sejam entrega da fatura, reaviso de vencimento, negatização, visita de cobrança, corte e assessoria de cobrança. A indisponibilidade de informações na base de dados sobre o comportamento de pagamento, somente foi superada através da geração de bases extras através de *queries* submetidas à base do sistema comercial. A identificação da efetividade dos meios de cobrança possui um duplo objetivo. Por um lado, possibilitar a otimização e priorização das ações de cobrança, minimizando os custos operacionais; por outro, servir como *input* fundamental na formação de um *score* de comportamento de pagamento, o que possibilita minimizar a inadimplência, através da priorização dos consumidores com maior probabilidade de inadimplência.

A medida numérica do risco de inadimplência foi obtida através de um modelo de regressão logística binária, que se baseou em informações estritamente objetivas disponíveis na base do sistema comercial da concessionária. Os resultados apresentados mostraram-se satisfatórios quanto ao poder de previsão do risco de inadimplência ou poder de classificação dos consumidores. Ao todo, 84,0% dos consumidores foram classificados corretamente, resultado considerado satisfatório. O modelo se mostrou mais eficaz para classificar consumidores inadimplentes do que adimplentes. Portanto, também é uma característica positiva tendo em vista que o objetivo primário desse tipo de modelo é prever a inadimplência. Assim, o modelo de regressão logística foi eficaz na medição da incerteza quanto ao pagamento das faturas mensais de energia elétrica.

Os estudos apresentados nesse trabalho confirmam a contribuição dos modelos de análise de agrupamentos e de medição do risco de inadimplência na gestão da inadimplência.

Os resultados obtidos indicam que o modelo é bastante eficiente em identificar em massa clientes que possivelmente se tornariam inadimplentes. Assim, ao programar uma ação de cobrança ou simplesmente de relacionamento e marketing, com o objetivo de minimizar os custos com inadimplência, as informações obtidas pelo modelo propiciarão que se atinjam grupos específicos e priorizar os casos com maior risco de inadimplência. Portanto, com a adoção dessas técnicas, espera-se uma melhoria no resultado operacional da concessionária com reflexos para a modicidade tarifária.

6.2 RECOMENDAÇÕES

Os resultados obtidos indicam algumas ações que podem contribuir para melhorar a eficiência no uso dos recursos disponíveis para a cobrança e reduzir a inadimplência dos consumidores.

A Coelba poderia utilizar na seleção de carteiras de cobrança, a segmentação de cobrança, os perfis de comportamento de pagamento e o modelo logístico de previsão de inadimplência para melhorar a efetividade das ações de cobrança.

Quanto aos perfis de comportamento de pagamento apresentados, recomenda-se a avaliação mais detalhada desses perfis. A obtenção de informações detalhadas, especialmente dos bons pagadores, pode propiciar ações de estímulo ao pagamento da fatura até a data de vencimento, o que evitaria custos com cobrança.

Ao modelo logístico de previsão de inadimplência, pode ser adicionado um modelo de cálculo do impacto do consumidor inadimplente nos objetivos financeiros da concessionária (severidade da inadimplência) com o objetivo de melhorar a eficiência na priorização dos consumidores.

6.3 DESENVOLVIMENTO DE TRABALHOS FUTUROS

Um desafio recente para as concessionárias de distribuição é o de aprofundar o entendimento sobre o comportamento de seus consumidores. Portanto, devem construir e manter uma base de dados que registre informações detalhadas sobre o consumidor e o relacionamento com os produtos e serviços.

A utilização de técnicas de otimização através da utilização de programação linear podem indicar novas pesquisas com o objetivo de melhorar a utilização dos recursos disponíveis e minimizar a inadimplência.

Novos trabalhos podem ser desenvolvidos incluindo variáveis que não foram consideradas neste estudo porque não estavam disponíveis na base de dados da empresa. Por exemplo, variáveis como histórico de reclamações, ocupação, renda, faixa etária, enfim, essas variáveis poderiam definir melhor o perfil dos consumidores.

Por fim, esse trabalho também pode ser desenvolvido utilizando outras técnicas, a exemplo de Redes Neurais Artificiais (RNA) e Análise Envoltória de Dados (*Data Envelopment Analysis (DEA)*).

REFERÊNCIAS

ABRADEE - ASSOCIAÇÃO BRASILEIRA DE DISTRIBUIDORAS DE ENERGIA ELÉTRICA. [Site Oficial]. Disponível em: <www.abradee.org.br> Acesso em: 25 jan. 2007.

ANEEL - AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Contrato de concessão n.º. 010/1997 - Coelba**. Disponível em: <www.aneel.gov.br> Acesso em: 12 mar. 2007.

_____. **Manual de contabilidade do serviço público de energia elétrica - MCSPE**. Disponível em: <www.aneel.gov.br> Acesso em: 2 maio 2006.

_____. **Nota Técnica n.º. 030/2003-SRE/ANEEL**: informações sobre a metodologia e os conceitos adotados pela ANEEL para proceder a revisão tarifária periódica das concessionárias de distribuição de energia elétrica. Brasília, 2003b. Disponível em: <www.aneel.gov.br> Acesso em: 11 jan. 2007.

_____. **Nota Técnica n.º. 052/ 2003 - SRE / ANEEL**: revisão tarifária periódica da concessionária de distribuição de energia elétrica Companhia de Eletricidade do Estado da Bahia - COELBA. Brasília, 2003a. Disponível em: <www.aneel.gov.br> Acesso em: 10 jan. 2007.

_____. **Nota Técnica n.º. 262/SRE/SFF/SRD/SRC/SFE/ANEEL**: revisão tarifária periódica das concessionárias de serviço público de distribuição de energia elétrica - 2º ciclo. Disponível em: <www.aneel.gov.br> Acesso em: 19 nov. 2007.

_____. **Nota Técnica n.º. 326/2002-SRE/ANEEL**: cálculo do fator X na revisão tarifária periódica das concessionárias de distribuição de energia elétrica: proposta metodológica. Brasília, 2002. Disponível em: <www.aneel.gov.br> Acesso em: 12 mar. 2006.

_____. **Nota Técnica n.º. 349/ 2007 - SRE / ANEEL**: metodologia de tratamento regulatório para receitas irrecuperáveis para o segundo ciclo de revisão tarifária periódica das concessionárias de distribuição de energia elétrica. Brasília, 2007. Disponível em: <www.aneel.gov.br> Acesso em: 15 jan. 2008.

_____. **Resolução Homologatória n.º. 638, de 17/04/2008**. Brasília, 2008. Disponível em: <www.aneel.gov.br> Acesso em: 22 jun. 2008.

_____. **Resolução Normativa n.º. 234, de 31/10/2006**: estabelece os conceitos gerais, as metodologias aplicáveis e os procedimentos iniciais para realização do segundo ciclo de Revisão Tarifária Periódica das concessionárias de serviço público de distribuição de energia elétrica. Disponível em: <www.aneel.gov.br> Acesso em: 15. nov. 2007.

_____. **Resolução Normativa n.º. 444, de 26/10/2001**: institui o Manual de Contabilidade do Serviço Público de Energia Elétrica, englobando o Plano de Contas revisado, com instruções contábeis e roteiro para elaboração e divulgação de informações econômicas e financeiras. Brasília, 2002. Disponível em: <www.aneel.gov.br> Acesso em: 14 maio 2006.

_____. **Resolução Normativa nº.456, de 29/11/2000:** estabelece as disposições atualizadas e consolidadas, relativas às condições gerais de fornecimento de energia elétrica, a serem observadas na prestação e utilização do serviço público de energia elétrica, tanto pelas concessionárias e permissionárias, quanto pelos consumidores. Brasília, 2000. Disponível em: <www.aneel.gov.br> Acesso em: 25 maio 2006.

ARAUJO, E. A. ; CARMONA, C. U. DE M. ; DEBOÇÃ, L. P. Risco de Crédito: construção de modelos Credit Scoring com abordagem de regressão logística para a análise da inadimplência de uma instituição de micro-crédito. In: SEMEAD - SEMINÁRIOS EM ADMINISTRAÇÃO FEA/USP, 9., São Paulo, **Anais eletrônicos....** São Paulo, 2006. Disponível em: <http://www.ead.fea.usp.br/semead>. Acesso em: 14 set. de 2006.

ARAÚJO, J. L. R. H. DE. **Fundamentos econômicos da regulação.** Rio de Janeiro: [S.n.],2001. Apresentação em formato PDF.

_____. Regulação de monopólios e mercados: questões básicas. In: SEMINÁRIO NACIONAL DO NÚCLEO DE ECONOMIA DA INFRA-ESTRUTURA, 1., 1997, São Paulo. **Anais ...** São Paulo, 1997.

BANCO CENTRAL. [Site oficial]. Disponível em: <www.bnb.gov.br> Acesso em: 25 maio 2008.

BARTH, N. L. Análise quantitativa para previsão de inadimplência. In: CONGRESSO INTERNACIONAL DE TECNOLOGIA DA INFORMAÇÃO - CATI. 2004, São Paulo. **Anais eletrônicos...** São Paulo: FGV/EASP, 2004. Disponível em: <<http://www.fgvsp.br/cati2004>>. Acesso em: 18 mar. 2005.

BERNSTEIN, P. L. **Desafio aos deuses:** a fascinante história do risco. 16. ed. Rio de Janeiro: Elsevier Editora, 1997.

BOGDANSKI, J.; TOMBINI, A. A. ; WERLANG, S.R. da C. **Implementing inflation targeting in Brazil.** Brasília: Banco Central do Brasil, 2000. (Trabalhos para Discussão, 1).

BOSNIA, ANTÔNIO CEZAR. **Mensuração das perdas dos processos produtivos: uma abordagem metodológica de controle interno.** Florianópolis, 1995. 106 f. (Tese de doutorado)- Universidade Federal de Santa Catarina -U FSC; Programa de pós-graduação em engenharia da produção, 1995.

BRASIL. Lei 8.897 de 13/02/1995, republicada em 29/08/1998. Dispõe sobre o regime de concessão e permissão da prestação de serviços públicos previstos no art. 175 da Constituição Federal de 1988. Brasília: **Diário Oficial da União**, 29 ago. 1998.

_____. Lei nº. 9.074 de 07/07/1995, republicada em 28/08/1998. Estabelece as normas para outorga e prorrogações das concessões e permissões da prestação de serviço público e dá outras providências. Brasília: **Diário Oficial da União**, 29 ago. 1998.

_____. Ministério das Minas e Energia - MME. **Modelo institucional do setor elétrico**. Brasília, 2003.

BRASILIANO, A. C. R. Riscos corporativos: novo enfoque na competitividade nas empresas. **Revista Eletrônica KPMG**, São Paulo, n. 1, p.21, mar./abr. 2002.

Disponível em: <<http://www.kpmg.com.br>>. Acesso em: 12 dez. 2004.

CAOINETTE, J. ; ALTMAN, E. ; NARAYANAN, P. **Gestão do risco de crédito: o próximo grande desafio financeiro**. Rio de Janeiro: Qualitymark, 1999. 500 p. (Serasa - Dinâmica do conhecimento).

CASTRO JUNIOR, FRANCISCO H. F. DE. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**. São Paulo, 2003. 169 f. (Dissertação de Mestrado) - Universidade de São Paulo - USP; FEA/USP, 2003.

COELBA - COMPANHIA DE ELETRICIDADE DO ESTADO DA BAHIA. **Análise dos resultados das pesquisas ANEEL de satisfação de clientes residenciais: 2000, 2001, 2002, 2003**. Salvador: COELBA, 2004. 13 p.

_____. **Painel Brasil. Pesquisa Inadimplência Coelba**. Salvador, 2002. 40 p.

_____. **Relatório anual 2004**. Salvador, 2005. 103 p.

_____. **Relatório anual 2005**. Salvador, 2006. 61 p.

COOPERS & LIBRAND. **Projeto RESEB - Relatório consolidado**. Rio de Janeiro: Eletrobrás, 1997. 39 p.

CROUHY, M.; GALAI, D. MARK, R. **Gerenciamento de risco: uma abordagem conceitual e prática: visão integrada dos riscos de crédito, operacional e de mercado**. Rio de Janeiro: Qualitymark, 2004. 664 p. (Serasa - Dinâmica do conhecimento).

DUKE ENERGY BRASIL. **Dicionário de energia elétrica inglês-português: terminologia do setor elétrico brasileiro**. São Paulo: Duke-energy, 2003. 133 p.

EIFERT, D. S. **Análise quantitativa na concessão de crédito versus inadimplência: um estudo empírico**. 2003. 81 f. (Dissertação)- Universidade Federal do Rio Grande do Sul - Escola de Administração - Programa de Pós Graduação em Administração, 2003.

FAIR ISAAC CORPORATION. **Understanding your credit score**. Disponível em: <www.fairisaac.com> Acesso em: 15 jul. 2005.

FERREIRA, D. F. **Estatística multivariada**. Lavras: Universidade Federal de Lavras, 1996. 400 f.

GREENACRE, M. **From correspondence analysis to multiple joint correspondence**. Barcelona: Universidad Pompeu Fabra, 1994. 42 p.

HAIR, J. F. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005. 593 p.

IGUTI, M. H. **Credit scoring**: um comparativo entre regressão logística e redes neurais. São Paulo: Universidade de São Paulo - USP, 2005.

JANOT, M. M. **Modelos de previsão de insolvência bancária no Brasil**. Brasília: Banco Central do Brasil, 2001, p. 42 (Trabalhos para discussão, 13).

KRZANOWSKY, W. J. **Principles of multivariate analysis**. Oxford: Clarendon Press, 1988.

LAFFONT, J. J.; TIROLE, J. **A Theory of incentives in procurement and regulation**. 1. ed. Cambridge, MA: MIT Press, 1993. 731 p.

MARTÍN, M. L. R. **Perdidas de energia**. Montevideu/Uruguai: CIER - Comisión de Integración Eléctrica Regional, 1991. 158 p.

MILLER, ROGER LEROY. **Microeconomia: teoria, questões e aplicações**. São Paulo: McGraw Hill do Brasil, 1981. 507 p.

MOITA NETO, J. M. **Estatística multivariada**: uma visão didática-metodológica. Disponível em: <http://criticanarede.com/cien_estatistica.html. Acesso em: 12 mar. 2006.

OHLSON, J. A. Financial ratios and the probabilistic prediction of bankruptcy. **Journal of Accounting Research**, v. 18, n. 1, primavera 1980.

PEREIRA, S. L. G. Na mira do crédito. **Revista GV Executivo**, São Paulo Fundação Getúlio Vargas (FGVSP), v. 5, n. 1, p. 31-36, fev./abr. 2006.

PERERA, L. C. J. Utilização da cadeia de Markov para avaliação de carteiras de crédito. In: SEMEAD - SEMINÁRIOS EM ADMINISTRAÇÃO FEA/USP, 3., 1998, São Paulo. **Anais eletrônicos...** São Paulo: FEA/USP, 1998. Disponível em: <http://www.ead.fea.usp.br/semead>. Acesso em: 17 maio 2006.

PEROBELLI, F. F. C. **Um modelo para gerenciamento de riscos em instituições não financeiras**: aplicação ao setor de distribuição de energia elétrica no Brasil. 147 f. 2004. Tese (Doutorado)- Universidade de São Paulo - USP; FEA/USP, 2004.

PINHEIRO, A. C. ; CABRAL, C. Mercado de crédito no Brasil: o papel do judiciário e de outras instituições. **Ensaio BNDES**, Rio de Janeiro, n.9, 1998. 87 p.

PINTO JUNIOR, H. Q. ; SILVEIRA, J. **Aspectos técnicos de regulação econômica: controle de preços**. Brasília: ANP - Agência Nacional do Petróleo, 1999.

PIRES, J. C. L. ; PICININI, M. S. **Mecanismos de regulação tarifária**: a experiência internacional e o caso brasileiro. Rio de Janeiro: BNDES, p. 57, 1998. (Textos para discussão, 64).

POSSAS, M. L. ; PONDÉ, J. L. ; FAGUNDES, J. **Regulação da Concorrência nos Setores de Infraestrutura no Brasil**: elementos para um quadro conceitual. Rio de Janeiro: UFRJ. Rio de Janeiro, p. 44, 2004.

PRESS, J. ; WILSON, S. Choosing between logistic regression and discriminant analysis. **Journal of the American Statistical Association**, v. 73, n. 364, p. 699-705, 1978.

PRICEWATERHOUSE & COOPERS. **Estudo do novo modelo do setor elétrico brasileiro**. São Paulo: P&C, 2004. 51 p.

QUEIROZ, R.S. B. A importância dos Modelos de Credit Scoring na concessão de crédito ao consumidor no varejo. In: SEMEAD - SEMINÁRIOS EM ADMINISTRAÇÃO, FEA/USP. 9., 2006, São Paulo, **Anais eletrônicos...** São Paulo, 2006. Disponível em: <<http://www.ead.fea.usp.br/semead>>. Acesso em: 12 set. 2006.

QUINECHE REYNA, F. **Gerenciamento de riscos I**. São Paulo: Instituto de Matemática Pura e Aplicada (IMPA), 2004. Slides apresentado no Mestrado em Energia.

SAMP - SEMINÁRIO ABRADÉE DE MELHORES PRÁTICAS – GESTÃO OPERACIONAL: GESTÃO DA INADIMPLÊNCIA, 2005. **Anais...** Campinas, SP: ABRADÉE, 2005.

SAUNDERS, A. **Medindo o risco de crédito**: novas abordagens para o value at risk e outros paradigmas. Rio de Janeiro: Qualitymark, 2000. 200 p. (Serasa - Dinâmica do conhecimento).

SECURATO, J. R. Modelo para determinação do valor presente de uma carteira de crédito e de seu risco: caso C.D.C. In: SEMEAD - SEMINÁRIOS EM ADMINISTRAÇÃO FEA/USP, 3., 1998, São Paulo. **Anais eletrônicos...** São Paulo: FEA/USP, 1998. Disponível em: <http://www.ead.fea.usp.br/semead>. Acesso em: 17 maio 2006.

SEN, A. K. **Desenvolvimento como liberdade**. São Paulo: Companhia das Letras, 2000. 409 p. (Prêmio Nobel - Companhia das Letras).

SERASA. [Site oficial]. Disponível em: <www.serasa.com.br> Acesso em: 25 maio 2008.

SILVA, R. K. M., CUNHA, F. M. C. **Projeto de inteligência corporativa para redução de inadimplência e perdas não-técnicas na Coelba**. Salvador: Coelba, 2004. 56 p.

SPSS INC. **SPSS® Base 13.0 user's guide** Chicago/USA, 2005. 744 p.

_____. **SPSS Regression Models™ 13.0** Chicago/USA, 2004. 70 p.

SPSS INC.; MEUMAN, J. J. ; HEISER, W. J. **SPSS Categories® 13.0** Chicago/USA: SPSS Inc, 2004. 371 p.

STANDARD & POOR'S. **Industry report card: latin american electric utilities**. 2004. Disponível em: <ratingsdirect.com>. Acesso em: 16 abr. 2006.

STATSOFT INC. **Correspondence analysis**. Disponível em: <statsoft.com> Acesso em: 18 out. 2006.

STIGLITZ, J. E.; WALSH, C. E. **Introdução a microeconomia**. 3. ed. Rio de Janeiro: Campus, 2003. 387 p.

TANURE, J. E. P. S. **Proposta de procedimentos e metodologia para estabelecimento de metas de qualidade (DEC e FEC) para concessionárias de distribuição de energia elétrica através da análise comparativa**. São Paulo, 2004. 249 f. Tese - Departamento de engenharia e automação, USP - Universidade de São Paulo.

THOMAS, L. C. A survey of credit and behavioural scoring: forecasting financial risk of lending consumers. **International Journal of Forecasting**, Edimburg - United Kingdon, v. 16, p. 149-172, 2000.

TORRES, A.; GREENACRE, M. **Measuring asymmetries in brand associations using correspondence analysis**. Barcelona/Espanha: Universidad Pompeu Fabra, 2003. 26 p.

TRIVELLONI, C. A. P.; HOCHHEIM, N. Avaliação em massa de imóveis por inferência estatística e análise multivariada. In: CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS (X-COBREAP). 10., 1999, Porto Alegre, **Anais...** Porto Alegre: IBAPE - Instituto Brasileiro de Avaliações e Perícias, 1999.

VARIAN, H. R. **Microeconomia: princípios básicos**. 4 ed. Rio de Janeiro: Campus, 1994. 709 p.

VASCONCELOS, M. S. de. **Proposta de método para análise de concessões de crédito a pessoas físicas**. São Paulo, 2002. 119 f. Dissertação de mestrado - FEA/USP, USP - Universidade de São Paulo.

VILELLA, V. A.; MACIEL, C. S. **A regulação do setor de infra-estrutura econômica: uma comparação internacional**. Brasília: IPEA, 1999. 112 p. (Textos para discussão, 684).

WILSON, N.; SUMMERS, B.; HOPE, R. Using payment behaviour data for credit risk modeling. **International Journal of the Economics and Business**, v. 7, n. 3, p. 338-349, 2001.

ZAMBALDI, F. et al. A questão da seleção adversa no microcrédito produtivo orientado: um estudo empírico sobre o comportamento de adimplência de empreendedores de baixa renda. **Pesquisa & Debate**, São Paulo, p. 309-331, nov. 2005.

ANEXO A – Esquema de aglomeração de análise hierárquica pelo método centróide

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	792	800	1,0000	0	0	406
2	798	799	1,0000	0	0	403
3	796	797	1,0000	0	0	404
4	794	795	1,0000	0	0	405
5	791	793	1,0000	0	0	407
6	784	790	1,0000	0	0	410
7	788	789	1,0000	0	0	408
8	786	787	1,0000	0	0	409
9	779	785	1,0000	0	0	413
10	782	783	1,0000	0	0	411
11	780	781	1,0000	0	0	412
12	774	778	1,0000	0	0	414
13	771	777	1,0000	0	0	415
14	775	776	1,0000	0	0	399
15	766	773	1,0000	0	0	418
16	768	772	1,0000	0	0	417
17	769	770	1,0000	0	0	416
18	617	767	1,0000	0	0	401
19	764	765	1,0000	0	0	419
20	761	763	1,0000	0	0	420
21	759	762	1,0000	0	0	421
22	756	760	1,0000	0	0	422
23	754	758	1,0000	0	0	423
24	753	757	1,0000	0	0	424
25	750	755	1,0000	0	0	426
26	751	752	1,0000	0	0	425
27	744	749	1,0000	0	0	428
28	725	748	1,0000	0	0	436
29	746	747	1,0000	0	0	427
30	689	745	1,0000	0	0	452
31	742	743	1,0000	0	0	429
32	739	741	1,0000	0	0	430
33	733	740	1,0000	0	0	432
34	731	738	1,0000	0	0	433
35	734	737	1,0000	0	0	431
36	735	736	1,0000	0	0	400
37	718	732	1,0000	0	0	440
38	726	730	1,0000	0	0	435
39	728	729	1,0000	0	0	434
40	723	727	1,0000	0	0	437
41	699	724	1,0000	0	0	436
42	721	722	1,0000	0	0	438
43	719	720	1,0000	0	0	439
44	716	717	1,0000	0	0	441
45	712	715	1,0000	0	0	443
46	713	714	1,0000	0	0	442
47	663	711	1,0000	0	0	465
48	702	710	1,0000	0	0	447
49	708	709	1,0000	0	0	444
50	706	707	1,0000	0	0	445
51	704	705	1,0000	0	0	446
52	701	703	1,0000	0	0	448
53	693	700	1,0000	0	0	451
54	697	698	1,0000	0	0	449
55	695	696	1,0000	0	0	450
56	623	694	1,0000	0	0	486
57	687	692	1,0000	0	0	454
58	688	691	1,0000	0	0	453
59	686	690	1,0000	0	0	455
60	684	685	1,0000	0	0	456
61	680	683	1,0000	0	0	458
62	681	682	1,0000	0	0	457
63	678	679	1,0000	0	0	459
64	674	677	1,0000	0	0	461
65	675	676	1,0000	0	0	460
66	670	673	1,0000	0	0	462
67	572	672	1,0000	0	0	510
68	658	671	1,0000	0	0	469
69	668	669	1,0000	0	0	463
70	666	667	1,0000	0	0	464
71	662	665	1,0000	0	0	466
72	660	664	1,0000	0	0	467
73	659	661	1,0000	0	0	468
74	652	657	1,0000	0	0	471
75	654	655	1,0000	0	0	470
76	651	653	1,0000	0	0	472
77	644	650	1,0000	0	0	475
78	645	649	1,0000	0	0	474
79	647	648	1,0000	0	0	473
80	616	646	1,0000	0	0	489
81	642	643	1,0000	0	0	476
82	639	641	1,0000	0	0	477
83	636	640	1,0000	0	0	479
84	637	638	1,0000	0	0	478
85	634	635	1,0000	0	0	480
86	628	633	1,0000	0	0	483
87	631	632	1,0000	0	0	481
88	629	630	1,0000	0	0	482
89	626	627	1,0000	0	0	484
90	624	625	1,0000	0	0	485
91	620	622	1,0000	0	0	487
92	619	621	1,0000	0	0	488
93	614	618	1,0000	0	0	490
94	593	615	1,0000	0	0	440
95	562	613	1,0000	0	0	515
96	611	612	1,0000	0	0	491
97	600	610	1,0000	0	0	496
98	608	609	1,0000	0	0	492
99	606	607	1,0000	0	0	493
100	604	605	1,0000	0	0	494
101	602	603	1,0000	0	0	495
102	598	601	1,0000	0	0	497
103	597	599	1,0000	0	0	498
104	592	596	1,0000	0	0	500
105	594	595	1,0000	0	0	499
106	588	591	1,0000	0	0	501
107	586	590	1,0000	0	0	502
108	575	589	1,0000	0	0	508
109	585	587	1,0000	0	0	503
110	579	584	1,0000	0	0	506
111	582	583	1,0000	0	0	504
112	580	581	1,0000	0	0	505
113	570	578	1,0000	0	0	511
114	576	577	1,0000	0	0	507
115	573	574	1,0000	0	0	509
116	407	571	1,0000	0	0	579
117	561	569	1,0000	0	0	516
118	567	568	1,0000	0	0	512
119	565	566	1,0000	0	0	513
120	563	564	1,0000	0	0	514
121	559	560	1,0000	0	0	517
122	557	558	1,0000	0	0	518

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
124	554	555	1,0000	0	0	519
125	551	552	1,0000	0	0	521
126	503	550	1,0000	0	0	542
127	421	549	1,0000	0	0	577
128	547	548	1,0000	0	0	522
129	545	546	1,0000	0	0	523
130	541	544	1,0000	0	0	525
131	542	543	1,0000	0	0	524
132	539	540	1,0000	0	0	526
133	536	538	1,0000	0	0	527
134	528	537	1,0000	0	0	529
135	534	535	1,0000	0	0	432
136	530	533	1,0000	0	0	528
137	531	532	1,0000	0	0	698
138	527	529	1,0000	0	0	530
139	525	526	1,0000	0	0	531
140	523	524	1,0000	0	0	532
141	521	522	1,0000	0	0	533
142	510	520	1,0000	0	0	538
143	518	519	1,0000	0	0	534
144	516	517	1,0000	0	0	535
145	514	515	1,0000	0	0	536
146	460	513	1,0000	0	0	425
147	511	512	1,0000	0	0	537
148	507	509	1,0000	0	0	539
149	504	508	1,0000	0	0	541
150	505	506	1,0000	0	0	540
151	501	502	1,0000	0	0	543
152	499	500	1,0000	0	0	544
153	497	498	1,0000	0	0	545
154	494	496	1,0000	0	0	546
155	492	495	1,0000	0	0	547
156	487	493	1,0000	0	0	550
157	490	491	1,0000	0	0	548
158	488	489	1,0000	0	0	549
159	485	486	1,0000	0	0	551
160	481	484	1,0000	0	0	552
161	467	483	1,0000	0	0	559
162	477	482	1,0000	0	0	554
163	479	480	1,0000	0	0	553
164	473	478	1,0000	0	0	556
165	475	476	1,0000	0	0	555
166	471	474	1,0000	0	0	557
167	470	472	1,0000	0	0	558
168	462	469	1,0000	0	0	562
169	431	468	1,0000	0	0	572
170	463	466	1,0000	0	0	561
171	464	465	1,0000	0	0	560
172	459	461	1,0000	0	0	563
173	457	458	1,0000	0	0	564
174	454	456	1,0000	0	0	565
175	452	453	1,0000	0	0	566
176	448	451	1,0000	0	0	568
177	449	450	1,0000	0	0	567
178	446	447	1,0000	0	0	598
179	444	445	1,0000	0	0	568
180	438	443	1,0000	0	0	563
181	1	442	1,0000	0	0	409
182	440	441	1,0000	0	0	569
183	414	439	1,0000	0	0	597
184	428	437	1,0000	0	0	560

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
186	433	434	1,0000	0	0	571
187	427	432	1,0000	0	0	574
188	429	430	1,0000	0	0	573
189	424	426	1,0000	0	0	575
190	420	425	1,0000	0	0	558
191	422	423	1,0000	0	0	576
192	417	419	1,0000	0	0	555
193	416	418	1,0000	0	0	578
194	376	415	1,0000	0	0	578
195	412	413	1,0000	0	0	562
196	410	411	1,0000	0	0	554
197	408	409	1,0000	0	0	550
198	382	406	1,0000	0	0	557
199	400	405	1,0000	0	0	582
200	403	404	1,0000	0	0	580
201	401	402	1,0000	0	0	581
202	398	399	1,0000	0	0	583
203	396	397	1,0000	0	0	584
204	388	395	1,0000	0	0	585
205	393	394	1,0000	0	0	546
206	390	392	1,0000	0	0	545
207	386	389	1,0000	0	0	544
208	385	387	1,0000	0	0	586
209	383	384	1,0000	0	0	587
210	365	381	1,0000	0	0	561
211	377	380	1,0000	0	0	551
212	378	379	1,0000	0	0	588
213	374	375	1,0000	0	0	589
214	371	373	1,0000	0	0	590
215	352	372	1,0000	0	0	577
216	369	370	1,0000	0	0	591
217	367	368	1,0000	0	0	592
218	363	366	1,0000	0	0	592
219	330	364	1,0000	0	0	559
220	361	362	1,0000	0	0	591
221	357	360	1,0000	0	0	543
222	358	359	1,0000	0	0	590
223	354	356	1,0000	0	0	541
224	353	355	1,0000	0	0	589
225	350	351	1,0000	0	0	538
226	347	349	1,0000	0	0	533
227	345	348	1,0000	0	0	588
228	344	346	1,0000	0	0	530
229	342	343	1,0000	0	0	587
230	339	341	1,0000	0	0	528
231	338	340	1,0000	0	0	586
232	334	337	1,0000	0	0	527
233	325	336	1,0000	0	0	585
234	327	335	1,0000	0	0	542
235	332	333	1,0000	0	0	524
236	323	331	1,0000	0	0	523
237	233	329	1,0000	0	0	515
238	326	328	1,0000	0	0	512
239	322	324	1,0000	0	0	584
240	320	321	1,0000	0	0	583
241	260	319	1,0000	0	0	573
242	314	318	1,0000	0	0	482
243	316	317	1,0000	0	0	581
244	311	315	1,0000	0	0	580
245	257	313	1,0000	0	0	469
246	309	312	1,0000	0	0	522

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
248	303	308	1,0000	0	0	521
249	306	307	1,0000	0	0	576
250	304	305	1,0000	0	0	575
251	286	302	1,0000	0	0	519
252	300	301	1,0000	0	0	574
253	298	299	1,0000	0	0	571
254	296	297	1,0000	0	0	593
255	294	295	1,0000	0	0	570
256	292	293	1,0000	0	0	569
257	287	291	1,0000	0	0	567
258	288	290	1,0000	0	0	594
259	265	289	1,0000	0	0	595
260	98	285	1,0000	0	0	695
261	281	284	1,0000	0	0	566
262	282	283	1,0000	0	0	582
263	279	280	1,0000	0	0	565
264	277	278	1,0000	0	0	564
265	275	276	1,0000	0	0	596
266	272	274	1,0000	0	0	556
267	268	273	1,0000	0	0	518
268	270	271	1,0000	0	0	553
269	267	269	1,0000	0	0	552
270	254	266	1,0000	0	0	517
271	263	264	1,0000	0	0	549
272	250	262	1,0000	0	0	594
273	258	259	1,0000	0	0	548
274	255	256	1,0000	0	0	547
275	237	253	1,0000	0	0	572
276	251	252	1,0000	0	0	540
277	248	249	1,0000	0	0	593
278	238	247	1,0000	0	0	489
279	245	246	1,0000	0	0	539
280	243	244	1,0000	0	0	537
281	241	242	1,0000	0	0	516
282	235	240	1,0000	0	0	511
283	231	239	1,0000	0	0	536
284	226	236	1,0000	0	0	486
285	177	234	1,0000	0	0	473
286	228	232	1,0000	0	0	505
287	223	230	1,0000	0	0	535
288	225	227	1,0000	0	0	504
289	211	224	1,0000	0	0	484
290	221	222	1,0000	0	0	534
291	219	220	1,0000	0	0	532
292	217	218	1,0000	0	0	531
293	213	216	1,0000	0	0	502
294	214	215	1,0000	0	0	529
295	209	212	1,0000	0	0	499
296	208	210	1,0000	0	0	526
297	205	207	1,0000	0	0	497
298	203	206	1,0000	0	0	525
299	140	204	1,0000	0	0	579
300	201	202	1,0000	0	0	495
301	199	200	1,0000	0	0	494
302	197	198	1,0000	0	0	493
303	195	196	1,0000	0	0	520
304	192	194	1,0000	0	0	492
305	188	193	1,0000	0	0	514
306	190	191	1,0000	0	0	491
307	186	187	1,0000	0	0	513
308	184	185	1,0000	0	0	488

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
310	181	182	1,0000	0	0	509
311	179	180	1,0000	0	0	485
312	174	178	1,0000	0	0	507
313	175	176	1,0000	0	0	478
314	171	173	1,0000	0	0	477
315	169	172	1,0000	0	0	506
316	166	170	1,0000	0	0	474
317	163	168	1,0000	0	0	503
318	159	167	1,0000	0	0	456
319	138	164	1,0000	0	0	595
320	157	162	1,0000	0	0	501
321	155	161	1,0000	0	0	471
322	151	160	1,0000	0	0	481
323	114	158	1,0000	0	0	403
324	149	156	1,0000	0	0	500
325	79	154	1,0000	0	0	455
326	152	153	1,0000	0	0	468
327	125	150	1,0000	0	0	465
328	143	148	1,0000	0	0	466
329	87	147	1,0000	0	0	427
330	145	146	1,0000	0	0	498
331	141	144	1,0000	0	0	496
332	139	142	1,0000	0	0	464
333	134	137	1,0000	0	0	463
334	135	136	1,0000	0	0	490
335	131	133	1,0000	0	0	462
336	129	132	1,0000	0	0	487
337	128	130	1,0000	0	0	461
338	123	127	1,0000	0	0	483
339	124	126	1,0000	0	0	459
340	116	122	1,0000	0	0	458
341	120	121	1,0000	0	0	480
342	118	119	1,0000	0	0	479
343	115	117	1,0000	0	0	476
344	99	113	1,0000	0	0	453
345	108	112	1,0000	0	0	475
346	109	111	1,0000	0	0	423
347	85	110	1,0000	0	0	510
348	106	107	1,0000	0	0	472
349	104	105	1,0000	0	0	402
350	101	102	1,0000	0	0	416
351	97	100	1,0000	0	0	470
352	95	96	1,0000	0	0	467
353	93	94	1,0000	0	0	460
354	91	92	1,0000	0	0	457
355	83	90	1,0000	0	0	454
356	88	89	1,0000	0	0	450
357	82	86	1,0000	0	0	424
358	75	84	1,0000	0	0	449
359	80	81	1,0000	0	0	451
360	37	78	1,0000	0	0	446
361	76	77	1,0000	0	0	448
362	72	74	1,0000	0	0	445
363	60	73	1,0000	0	0	447
364	70	71	1,0000	0	0	444
365	68	69	1,0000	0	0	443
366	66	67	1,0000	0	0	415
367	64	65	1,0000	0	0	438
368	31	63	1,0000	0	0	404
369	61	62	1,0000	0	0	437
370	56	59	1,0000	0	0	442

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
496	141	600	0,5000	331	97	680
497	205	598	0,5000	297	102	688
498	145	597	0,5000	330	103	681
499	209	594	0,5000	295	105	686
500	149	592	0,5000	324	104	682
501	157	588	0,5000	320	106	683
502	213	586	0,5000	293	107	684
503	163	585	0,5000	317	109	687
504	225	582	0,5000	288	111	678
505	228	580	0,5000	286	112	677
506	169	579	0,5000	315	110	689
507	174	576	0,5000	312	114	692
508	261	575	0,5000	247	108	662
509	181	573	0,5000	310	115	693
510	85	572	0,5000	347	67	668
511	235	570	0,5000	282	113	674
512	326	567	0,5000	238	118	636
513	186	565	0,5000	307	119	697
514	188	563	0,5000	305	120	693
515	233	562	0,5000	237	95	675
516	241	561	0,5000	281	117	672
517	254	559	0,5000	270	121	666
518	268	557	0,5000	267	122	659
519	286	554	0,5000	251	124	653
520	195	553	0,5000	303	123	692
521	303	551	0,5000	248	125	647
522	309	547	0,5000	246	128	644
523	323	545	0,5000	236	129	638
524	332	542	0,5000	235	131	633
525	203	541	0,5000	298	130	689
526	208	539	0,5000	296	132	687
527	334	536	0,5000	232	133	632
528	339	530	0,5000	230	136	630
529	214	528	0,5000	294	134	683
530	344	527	0,5000	228	138	628
531	217	525	0,5000	292	139	682
532	219	523	0,5000	291	140	681
533	347	521	0,5000	226	141	626
534	221	518	0,5000	290	143	680
535	223	516	0,5000	287	144	679
536	231	514	0,5000	283	145	676
537	243	511	0,5000	280	147	671
538	350	510	0,5000	225	142	625
539	245	507	0,5000	279	148	670
540	251	505	0,5000	276	150	667
541	354	504	0,5000	223	149	622
542	327	503	0,5000	234	126	635
543	357	501	0,5000	221	151	621
544	386	499	0,5000	207	152	614
545	390	497	0,5000	206	153	613
546	393	494	0,5000	205	154	612
547	255	492	0,5000	274	155	665
548	258	490	0,5000	273	157	664
549	263	488	0,5000	271	158	661
550	408	487	0,5000	197	156	611
551	377	485	0,5000	211	159	616
552	267	481	0,5000	269	160	660
553	270	479	0,5000	268	163	658
554	410	477	0,5000	196	162	610
555	417	475	0,5000	192	165	608
556	272	473	0,5000	266	164	657

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
558	420	470	0,5000	190	167	607
559	330	467	0,5000	219	161	634
560	428	464	0,5000	184	171	606
561	365	463	0,5000	210	170	617
562	412	462	0,5000	195	168	609
563	438	459	0,5000	180	172	605
564	277	457	0,5000	264	173	656
565	279	454	0,5000	263	174	655
566	281	452	0,5000	261	175	654
567	287	449	0,5000	257	177	652
568	444	448	0,5000	179	176	604
569	292	440	0,5000	256	182	651
570	294	435	0,5000	255	185	650
571	298	433	0,5000	253	186	649
572	237	431	0,5000	275	169	673
573	260	429	0,5000	241	188	663
574	300	427	0,5000	252	187	648
575	304	424	0,5000	250	189	646
576	306	422	0,5000	249	191	645
577	352	421	0,5000	215	127	624
578	376	416	0,5000	194	193	599
579	140	407	0,5000	299	116	669
580	311	403	0,5000	244	200	643
581	316	401	0,5000	243	201	641
582	282	400	0,5000	262	199	698
583	320	398	0,5000	240	202	640
584	322	396	0,5000	239	203	639
585	325	388	0,5000	233	204	637
586	338	385	0,5000	231	208	631
587	342	383	0,5000	229	209	629
588	345	378	0,5000	227	212	627
589	353	374	0,5000	224	213	623
590	358	371	0,5000	222	214	620
591	361	369	0,5000	220	216	619
592	363	367	0,5000	218	217	618
593	248	296	0,5000	277	254	669
594	250	288	0,5000	272	258	668
595	138	265	0,5000	319	259	601
596	53	275	0,4500	397	265	699
597	15	414	0,4167	400	183	700
598	4	446	0,3750	405	178	701
599	13	376	0,3250	398	578	702
600	656	699	0,2917	399	436	790
601	43	138	0,2917	401	595	749
602	1	593	0,2500	409	440	703
603	460	534	0,2500	425	432	703
604	8	444	0,2500	407	568	705
605	10	438	0,2500	408	563	706
606	12	428	0,2500	413	560	707
607	18	420	0,2500	417	558	709
608	22	417	0,2500	420	555	711
609	37	412	0,2500	446	562	710
610	24	410	0,2500	422	554	713
611	27	408	0,2500	426	550	715
612	32	393	0,2500	431	546	720
613	34	390	0,2500	435	545	721
614	39	386	0,2500	439	544	724
615	79	382	0,2500	455	557	718
616	159	377	0,2500	456	551	718
617	31	365	0,2500	404	561	696
618	2	363	0,2500	406	592	697

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
620	16	358	0,2500	411	590	712
621	52	357	0,2500	441	543	726
622	56	354	0,2500	442	541	728
623	19	353	0,2500	412	589	714
624	66	352	0,2500	415	577	738
625	60	350	0,2500	447	538	729
626	75	347	0,2500	449	533	731
627	25	345	0,2500	414	588	717
628	88	344	0,2500	450	530	732
629	29	342	0,2500	418	587	719
630	99	339	0,2500	453	528	736
631	36	338	0,2500	419	586	722
632	116	334	0,2500	458	527	736
633	124	332	0,2500	459	524	732
634	82	330	0,2500	424	559	716
635	177	327	0,2500	473	542	710
636	87	326	0,2500	427	512	745
637	41	325	0,2500	421	585	723
638	128	323	0,2500	461	523	731
639	44	322	0,2500	428	584	725
640	46	320	0,2500	429	583	727
641	48	316	0,2500	430	581	730
642	165	314	0,2500	452	482	716
643	51	311	0,2500	433	580	733
644	131	309	0,2500	462	522	729
645	57	306	0,2500	434	576	734
646	61	304	0,2500	437	575	735
647	134	303	0,2500	463	521	728
648	64	300	0,2500	438	574	737
649	68	298	0,2500	443	571	739
650	70	294	0,2500	444	570	740
651	72	292	0,2500	445	569	741
652	76	287	0,2500	448	567	742
653	139	286	0,2500	464	519	726
654	80	281	0,2500	451	566	743
655	83	279	0,2500	454	565	744
656	91	277	0,2500	457	564	744
657	93	272	0,2500	460	556	743
658	95	270	0,2500	467	553	742
659	143	268	0,2500	466	518	724
660	97	267	0,2500	470	552	741
661	106	263	0,2500	472	549	740
662	238	261	0,2500	489	508	700
663	114	260	0,2500	403	573	704
664	108	258	0,2500	475	548	739
665	115	255	0,2500	476	547	737
666	152	254	0,2500	468	517	721
667	118	251	0,2500	479	540	735
668	85	250	0,2500	510	594	746
669	140	248	0,2500	579	593	699
670	120	245	0,2500	480	539	734
671	123	243	0,2500	483	537	733
672	155	241	0,2500	471	516	720
673	226	237	0,2500	486	572	704
674	166	235	0,2500	474	511	715
675	101	233	0,2500	416	515	747
676	129	231	0,2500	487	536	730
677	171	228	0,2500	477	505	713
678	175	225	0,2500	478	504	711
679	135	223	0,2500	490	535	727
680	141	221	0,2500	496	534	725

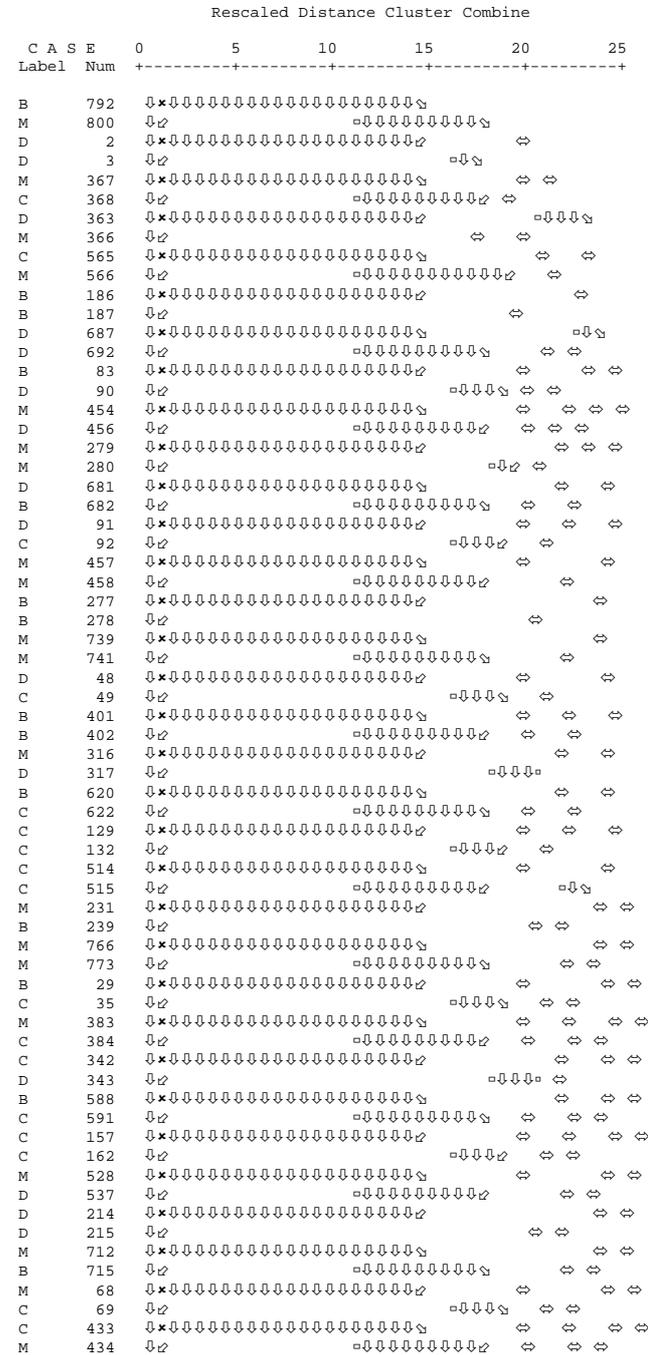
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
682	149	217	0,2500	500	531	722
683	157	214	0,2500	501	529	719
684	179	213	0,2500	485	502	709
685	109	211	0,2500	423	484	738
686	184	209	0,2500	488	499	707
687	163	208	0,2500	503	526	717
688	190	205	0,2500	491	497	706
689	169	203	0,2500	506	525	714
690	192	201	0,2500	492	495	705
691	197	199	0,2500	493	494	701
692	174	195	0,2500	507	520	712
693	181	188	0,2500	509	514	708
694	125	151	0,2500	465	481	702
695	98	103	0,2167	260	402	789
696	31	257	0,1875	617	469	745
697	2	186	0,1875	618	513	748
698	282	531	0,1750	582	137	789
699	53	140	0,1750	596	669	746
700	15	238	0,1625	597	662	750
701	4	197	0,1458	598	691	751
702	13	125	0,1403	599	694	747
703	1	460	0,1250	602	603	794
704	114	226	0,1250	663	673	793
705	8	192	0,1250	604	690	752
706	10	190	0,1250	605	688	755
707	12	184	0,1250	606	686	761
708	5	181	0,1250	619	693	753
709	18	179	0,1250	607	684	763
710	37	177	0,1250	609	635	771
711	22	175	0,1250	608	678	764
712	16	174	0,1250	620	692	754
713	24	171	0,1250	610	677	766
714	19	169	0,1250	623	689	756
715	27	166	0,1250	611	674	767
716	82	165	0,1250	634	642	768
717	25	163	0,1250	627	687	757
718	79	159	0,1250	615	616	750
719	29	157	0,1250	629	683	758
720	32	155	0,1250	612	672	772
721	34	152	0,1250	613	666	767
722	36	149	0,1250	631	682	759
723	41	145	0,1250	637	681	760
724	39	143	0,1250	614	659	766
725	44	141	0,1250	639	680	762
726	52	139	0,1250	621	653	764
727	46	135	0,1250	640	679	765
728	56	134	0,1250	622	647	763
729	60	131	0,1250	625	644	761
730	48	129	0,1250	641	676	770
731	75	128	0,1250	626	638	755
732	88	124	0,1250	628	633	752
733	51	123	0,1250	643	671	765
734	57	120	0,1250	645	670	762
735	61	118	0,1250	646	667	760
736	99	116	0,1250	630	632	751
737	64	115	0,1250	648	665	759
738	66	109	0,1250	624	685	769
739	68	108	0,1250	649	664	758
740	70	106	0,1250	650	661	757
741	72	97	0,1250	651	660	756
742	76	95	0,1250	652	658	754

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
743	80	93	0,1250	654	657	753
744	83	91	0,1250	655	656	748
745	31	87	0,1042	696	636	768
746	53	85	0,1028	699	668	792
747	13	101	0,0972	702	675	769
748	2	83	0,0729	697	744	770
749	43	229	0,0714	601	396	792
750	15	79	0,0697	700	718	771
751	4	99	0,0670	701	736	772
752	8	88	0,0625	705	732	777
753	5	80	0,0625	708	743	773
754	16	76	0,0625	712	742	774
755	10	75	0,0625	706	731	778
756	19	72	0,0625	714	741	775
757	25	70	0,0625	717	740	776
758	29	68	0,0625	719	739	780
759	36	64	0,0625	722	737	776
760	41	61	0,0625	723	735	775
761	12	60	0,0625	707	729	779
762	44	57	0,0625	725	734	774
763	18	56	0,0625	709	728	781
764	22	52	0,0625	711	726	779
765	46	51	0,0625	727	733	773
766	24	39	0,0625	713	724	778
767	27	34	0,0625	715	721	777
768	31	82	0,0563	745	716	790
769	13	66	0,0536	747	738	791
770	2	48	0,0491	748	730	780
771	15	37	0,0485	750	710	794
772	4	32	0,0479	751	720	781
773	5	46	0,0313	753	765	782
774	16	44	0,0313	754	762	783
775	19	41	0,0313	756	760	783
776	25	36	0,0313	757	759	782
777	8	27	0,0313	752	767	784
778	10	24	0,0313	755	766	785
779	12	22	0,0313	761	764	784
780	2	29	0,0270	770	758	786
781	4	18	0,0265	772	763	785
782	5	25	0,0156	773	776	788
783	16	19	0,0156	774	775	786
784	8	12	0,0156	777	779	787
785	4	10	0,0142	781	778	787
786	2	16	0,0105	780	783	788
787	4	8	0,0074	785	784	796
788	2	5	0,0064	786	782	795
789	98	282	-0,0709	695	698	791
790	31	656	-0,1147	768	600	795
791	13	98	-0,1147	769	789	793
792	43	53	-0,1323	749	746	797
793	13	114	-0,1530	791	704	798
794	1	15	-0,1576	703	771	799
795	2	31	-0,1922	788	790	796
796	2	4	-0,1920	795	787	797
797	2	43	-0,1969	796	792	798
798	2	13	-0,1994	797	793	799
799	1	2	-0,2328	794	798	0

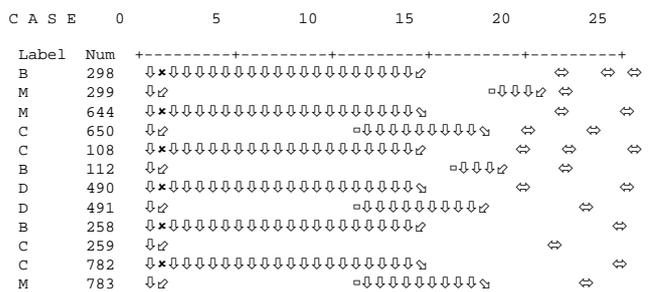
ANEXO B - Dendrograma de análise hierárquica pelo método centróide

***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****

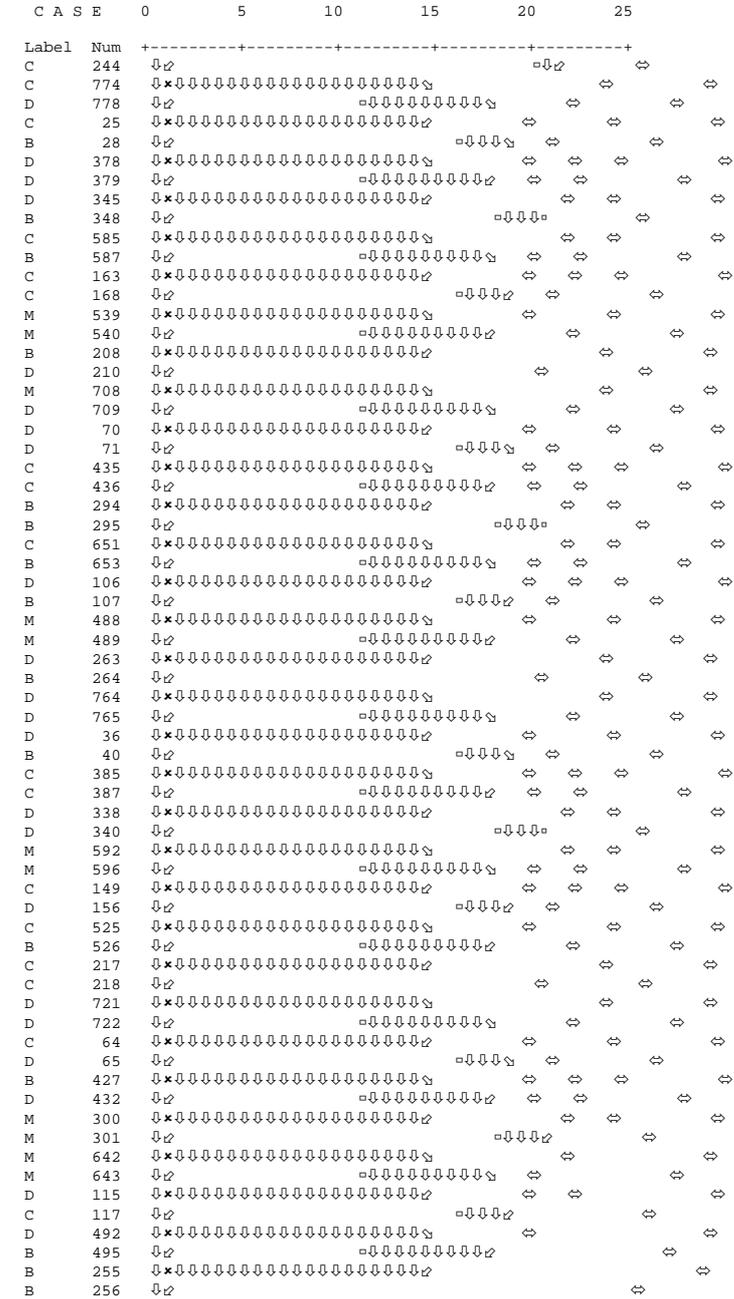
Dendrogram using Centroid Method



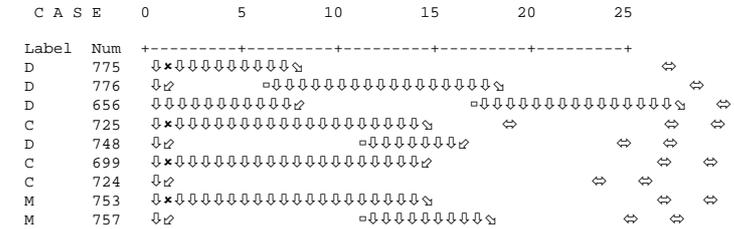
***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****




```
M 315  ↓↓
D 628  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 633  ↓↓
D 123  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 127  ↓↓
M 511  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 512  ↓↓
C 243  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
* * * * * H I E R A R C H I C A L C L U S T E R   A N A L Y S I S * * * * *
```



```
* * * * * H I E R A R C H I C A L C L U S T E R   A N A L Y S I S * * * * *
```




```

D 207  ↓↓
M 697  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 698  ↓↓
C 75   ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 84   ↓↓
D 521  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 522  ↓↓
C 347  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 349  ↓↓
M 674  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓

```

***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****

```

C A S E 0 5 10 15 20 25
Label Num +-----+-----+-----+-----+
C 677  ↓↓
M 128  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 130  ↓↓
M 545  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 546  ↓↓
B 323  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 331  ↓↓
M 756  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 760  ↓↓
B 24   ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 26   ↓↓
B 477  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 482  ↓↓
D 410  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 411  ↓↓
M 639  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 641  ↓↓
D 171  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 173  ↓↓
M 580  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 581  ↓↓
B 228  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 232  ↓↓
M 719  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 720  ↓↓
B 39   ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 50   ↓↓
D 499  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 500  ↓↓
D 386  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 389  ↓↓
D 662  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 665  ↓↓
M 143  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 148  ↓↓
D 557  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 558  ↓↓
M 268  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 273  ↓↓
D 768  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 772  ↓↓
C 18   ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 20   ↓↓
M 470  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 472  ↓↓
D 420  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 425  ↓↓
C 624  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 625  ↓↓
C 179  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 180  ↓↓
C 586  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 590  ↓↓
M 213  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 216  ↓↓
C 713  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 714  ↓↓
B 56   ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
B 59   ↓↓
M 504  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 508  ↓↓
C 354  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
C 356  ↓↓
M 668  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
M 669  ↓↓
M 134  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
D 137  ↓↓
M 551  ↓×↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓

```

***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****

```

C A S E 0 5 10 15 20 25
Label Num +-----+-----+-----+-----+
M 552  ↓↓

```


ANEXO C - Esquema de aglomeração de análise hierárquica pelo método de Ward

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1266	1774	0,0000	0	0	1576
2	1770	1773	0,0000	0	0	5
3	1743	1772	0,0000	0	0	31
4	1712	1771	0,0000	0	0	59
5	5	1770	0,0000	0	2	13
6	1765	1769	0,0000	0	0	10
7	1736	1768	0,0000	0	0	38
8	1727	1767	0,0000	0	0	46
9	1733	1766	0,0000	0	0	41
10	14	1765	0,0000	0	6	39
11	1762	1764	0,0000	0	0	13
12	635	1763	0,0000	0	0	1052
13	5	1762	0,0000	5	11	15
14	1760	1761	0,0000	0	0	15
15	5	1760	0,0000	13	14	19
16	1737	1759	0,0000	0	0	37
17	1756	1758	0,0000	0	0	19
18	1649	1757	0,0000	0	0	119
19	5	1756	0,0000	15	17	23
20	1752	1755	0,0000	0	0	23
21	1745	1754	0,0000	0	0	29
22	1508	1753	0,0000	0	0	250
23	5	1752	0,0000	19	20	36
24	1735	1751	0,0000	0	0	39
25	1575	1750	0,0000	0	0	186
26	1252	1749	0,0000	0	0	483
27	941	1748	0,0000	0	0	772
28	953	1747	0,0000	0	0	761
29	226	1745	0,0000	0	21	662
30	1738	1744	0,0000	0	0	36
31	140	1743	0,0000	0	3	554
32	1665	1742	0,0000	0	0	103
33	1716	1741	0,0000	0	0	56
34	852	1740	0,0000	0	0	855
35	1178	1739	0,0000	0	0	551
36	5	1738	0,0000	23	30	52
37	244	1737	0,0000	0	16	1273
38	344	1736	0,0000	0	7	118
39	14	1735	0,0000	10	24	108
40	1055	1734	0,0000	0	0	667
41	401	1733	0,0000	0	9	786
42	1338	1731	0,0000	0	0	407
43	1629	1730	0,0000	0	0	137
44	1641	1729	0,0000	0	0	126
45	1653	1728	0,0000	0	0	115
46	13	1727	0,0000	0	8	328
47	1720	1726	0,0000	0	0	52
48	1574	1725	0,0000	0	0	187
49	1381	1724	0,0000	0	0	368
50	1660	1723	0,0000	0	0	108
51	1159	1722	0,0000	0	0	570
52	5	1720	0,0000	36	47	61
53	1710	1719	0,0000	0	0	61
54	1587	1718	0,0000	0	0	176
55	1547	1717	0,0000	0	0	213
56	166	1716	0,0000	0	33	896
57	1611	1714	0,0000	0	0	154
58	115	1713	0,0000	0	0	1581
59	52	1712	0,0000	0	4	92
60	1141	1711	0,0000	0	0	587
61	5	1710	0,0000	52	53	67
62	1471	1709	0,0000	0	0	283
63	798	1707	0,0000	0	0	905
64	1703	1706	0,0000	0	0	67
65	1650	1705	0,0000	0	0	118
66	630	1704	0,0000	0	0	1057
67	5	1703	0,0000	61	64	75
68	1344	1702	0,0000	0	0	401
69	1309	1701	0,0000	0	0	433
70	1402	1700	0,0000	0	0	349
71	1695	1699	0,0000	0	0	75
72	1586	1698	0,0000	0	0	177
73	1406	1697	0,0000	0	0	345
74	1595	1696	0,0000	0	0	170
75	5	1695	0,0000	67	71	88
76	1682	1694	0,0000	0	0	88
77	1627	1693	0,0000	0	0	139
78	1655	1692	0,0000	0	0	113
79	1466	1691	0,0000	0	0	288
80	1677	1690	0,0000	0	0	92
81	1646	1689	0,0000	0	0	121
82	654	1688	0,0000	0	0	1033
83	1411	1687	0,0000	0	0	1570
84	1596	1686	0,0000	0	0	169
85	1651	1685	0,0000	0	0	117
86	1372	1684	0,0000	0	0	376
87	1533	1683	0,0000	0	0	227
88	5	1682	0,0000	75	76	94
89	1445	1681	0,0000	0	0	307
90	1041	1680	0,0000	0	0	678
91	1131	1678	0,0000	0	0	596
92	52	1677	0,0000	59	80	109
93	1675	1676	0,0000	0	0	94
94	5	1675	0,0000	88	93	96
95	1673	1674	0,0000	0	0	96
96	5	1673	0,0000	94	95	107
97	1669	1672	0,0000	0	0	100
98	1600	1671	0,0000	0	0	165
99	930	1670	0,0000	0	0	782
100	283	1669	0,0000	0	97	191
101	1171	1668	0,0000	0	0	558
102	1409	1667	0,0000	0	0	342
103	4	1665	0,0000	0	32	327
104	1659	1664	0,0000	0	0	109
105	1661	1663	0,0000	0	0	107
106	1513	1662	0,0000	0	0	246
107	5	1661	0,0000	96	105	123
108	14	1660	0,0000	39	50	160
109	52	1659	0,0000	92	104	208
110	1644	1658	0,0000	0	0	123
111	778	1657	0,0000	0	0	1545
112	1561	1656	0,0000	0	0	199
113	399	1655	0,0000	0	78	848
114	1632	1654	0,0000	0	0	134
115	131	1653	0,0000	0	45	762
116	1196	1652	0,0000	0	0	535
117	162	1651	0,0000	0	85	201
118	344	1650	0,0000	38	65	193
119	6	1649	0,0000	0	18	386
120	968	1648	0,0000	0	0	747
121	35	1646	0,0000	0	81	803
122	817	1645	0,0000	0	0	1580
123	5	1644	0,0000	107	110	129
124	1559	1643	0,0000	0	0	201
125	1422	1642	0,0000	0	0	330
126	383	1641	0,0000	0	44	861
127	1638	1640	0,0000	0	0	129
128	512	1639	0,0000	0	0	1555
129	5	1638	0,0000	123	127	131
130	1635	1637	0,0000	0	0	131
131	5	1635	0,0000	129	130	145
132	659	1634	0,0000	0	0	1578
133	1605	1633	0,0000	0	0	160
134	82	1632	0,0000	0	114	382
135	1421	1631	0,0000	0	896	331
136	1506	1630	0,0000	0	0	252
137	96	1629	0,0000	0	43	556
138	1425	1628	0,0000	0	0	327
139	132	1627	0,0000	0	77	683
140	1570	1626	0,0000	0	0	191
141	1289	1625	0,0000	0	0	452
142	1620	1624	0,0000	0	0	145
143	1619	1622	0,0000	0	0	146
144	1612	1621	0,0000	0	0	153
145	5	1620	0,0000	131	142	158
146	238	1619	0,0000	0	143	901
147	1493	1618	0,0000	0	0	1579
148	907	1617	0,0000	0	0	803
149	465	1616	0,0000	0	0	1201
150	1389	1615	0,0000	0	0	361
151	1461	1614	0,0000	0	0	1556
152	1360	1613	0,0000	0	0	386
153	1127	1612	0,0000	0	144	205
154	85	1611	0,0000	0	57	451
155	1607	1610	0,0000	0	0	158
156	1568	1609	0,0000	0	0	193
157	1006	1608	0,0000	0	0	710
158	5	1607	0,0000	145	155	162

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
319	1432	1433	0,0000	0	0	320
320	5	1432	0,0000	318	319	339
321	972	1431	0,0000	0	0	743
322	1399	1430	0,0000	0	0	352
323	1395	1429	0,0000	0	0	355
324	952	1428	0,0000	0	0	762
325	60	1427	0,0000	303	314	343
326	1413	1426	0,0000	0	0	339
327	4	1425	0,0000	103	138	406
328	13	1424	0,0000	46	230	591
329	515	1423	0,0000	0	0	1572
330	246	1422	0,0000	0	125	971
331	697	1421	0,0000	0	135	1607
332	98	1420	0,0000	0	224	369
333	136	1419	0,0000	0	0	1590
334	1380	1418	0,0000	0	0	369
335	1416	1417	0,0000	0	0	336
336	344	1416	0,0000	263	335	381
337	911	1415	0,0000	0	0	1581
338	1408	1414	0,0000	0	0	343
339	5	1413	0,0000	320	326	350
340	1401	1412	0,0000	0	0	350
341	52	1410	0,0000	261	280	405
342	175	1409	0,0000	0	102	835
343	60	1408	0,0000	325	338	426
344	14	1407	0,0000	306	308	377
345	254	1406	0,0000	0	73	410
346	1	1405	0,0000	0	249	394
347	1341	1404	0,0000	0	0	404
348	159	1403	0,0000	0	174	613
349	78	1402	0,0000	0	70	739
350	5	1401	0,0000	339	340	370
351	1371	1400	0,0000	0	0	377
352	22	1399	0,0000	299	322	387
353	1118	1398	0,0000	0	0	607
354	1074	1396	0,0000	0	0	649
355	199	1395	0,0000	213	323	488
356	1340	1394	0,0000	0	0	405
357	284	1393	0,0000	0	0	1595
358	91	1392	0,0000	0	189	759
359	1240	1391	0,0000	0	0	493
360	1339	1390	0,0000	0	0	406
361	71	1389	0,0000	0	150	1304
362	1167	1387	0,0000	0	0	562
363	1130	1386	0,0000	0	0	597
364	76	1385	0,0000	0	254	500
365	1379	1384	0,0000	0	0	370
366	1367	1383	0,0000	0	0	381
367	31	1382	0,0000	0	297	481
368	597	1381	0,0000	0	49	1023
369	98	1380	0,0000	332	334	408
370	5	1379	0,0000	350	365	391
371	746	1378	0,0000	246	285	900
372	1290	1377	0,0000	0	0	1550
373	1174	1375	0,0000	0	0	555
374	124	1374	0,0000	0	316	1075
375	540	1373	0,0000	238	298	956
376	214	1372	0,0000	0	86	1200
377	14	1371	0,0000	344	351	501
378	1359	1370	0,0000	0	0	387
379	1355	1369	0,0000	0	0	391
380	1337	1368	0,0000	0	0	408
381	344	1367	0,0000	336	366	520
382	82	1365	0,0000	134	219	399
383	806	1364	0,0000	0	0	898
384	1300	1362	0,0000	0	0	442
385	1351	1361	0,0000	0	0	394
386	6	1360	0,0000	119	152	557
387	22	1359	0,0000	352	378	428
388	922	1358	0,0000	0	0	789
389	1335	1357	0,0000	0	0	410
390	1314	1356	0,0000	0	300	1611
391	5	1355	0,0000	370	379	393
392	1353	1354	0,0000	0	0	393
393	5	1353	0,0000	391	392	402
394	1	1351	0,0000	346	385	668
395	287	1350	0,0000	0	0	1348
396	1061	1349	0,0000	0	0	662
397	1036	1348	0,0000	0	0	683
398	1346	1347	0,0000	0	0	399

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
399	82	1346	0,0000	382	398	476
400	1343	1345	0,0000	0	0	402
401	143	1344	0,0000	0	68	1243
402	5	1343	0,0000	393	400	415
403	1297	1342	0,0000	0	0	445
404	3	1341	0,0000	0	347	1326
405	52	1340	0,0000	341	356	436
406	4	1339	0,0000	327	360	1041
407	738	1338	0,0000	0	42	532
408	98	1337	0,0000	369	380	512
409	15	1336	0,0000	0	204	891
410	254	1335	0,0000	345	389	990
411	1220	1334	0,0000	0	0	512
412	1317	1332	0,0000	0	0	426
413	1053	1331	0,0000	0	0	1589
414	1329	1330	0,0000	0	0	415
415	5	1329	0,0000	402	414	421
416	1323	1328	0,0000	0	0	421
417	1315	1327	0,0000	0	0	428
418	1293	1326	0,0000	0	0	449
419	1254	1325	0,0000	0	0	481
420	609	1324	0,0000	0	0	1075
421	5	1323	0,0000	415	416	423
422	1320	1322	0,0000	0	0	423
423	5	1320	0,0000	421	422	446
424	816	1319	0,0000	0	0	888
425	1306	1318	0,0000	0	0	436
426	60	1317	0,0000	343	412	498
427	1241	1316	0,0000	0	0	1563
428	22	1315	0,0000	387	417	443
429	1160	1313	0,0000	0	0	569
430	1010	1312	0,0000	0	0	706
431	1233	1311	0,0000	0	0	500
432	27	1310	0,0000	266	291	549
433	39	1309	0,0000	0	69	629
434	1296	1308	0,0000	0	0	446
435	32	1307	0,0000	0	305	576
436	52	1306	0,0000	405	425	456
437	700	1305	0,0000	0	178	1604
438	1284	1304	0,0000	0	0	456
439	1299	1303	0,0000	0	0	443
440	1301	1302	0,0000	0	0	441
441	363	1301	0,0000	0	440	928
442	138	1300	0,0000	0	384	564
443	22	1299	0,0000	428	439	468
444	1165	1298	0,0000	0	0	564
445	279	1297	0,0000	0	403	1568
446	5	1296	0,0000	423	434	462
447	1181	1295	0,0000	0	0	549
448	97	1294	0,0000	227	294	497
449	422	1293	0,0000	211	418	571
450	1189	1292	0,0000	0	0	542
451	85	1291	0,0000	154	194	721
452	47	1289	0,0000	0	141	590
453	1136	1288	0,0000	0	0	591
454	1269	1287	0,0000	0	0	468
455	945	1286	0,0000	0	0	769
456	52	1284	0,0000	436	438	545
457	415	1283	0,0000	0	281	1189
458	89	1282	0,0000	0	223	718
459	1270	1281	0,0000	0	0	467
460	1277	1280	0,0000	0	0	462
461	1259	1278	0,0000	0	0	476
462	5	1277	0,0000	446	460	470
463	1212	1276	0,0000	0	0	520
464	1265	1275	0,0000	0	0	470
465	1173	1274	0,0000	0	0	556
466	1088	1271	0,0000	0	0	636
467	271	1270	0,0000	292	459	628
468	22	1269	0,0000	443	454	533
469	1096	1267	0,0000	0	0	628
470	5	1265	0,0000	462	464	477
471	70	1264	0,0000	0	0	1578
472	1258	1263	0,0000	0	0	477
473	1122	1262	0,0000	0	0	604
474	976	1261	0,0000	0	0	739
475	1247	1260	0,0000	0	0	488
476	82	1259	0,0000	399	461	522
477	5	1258	0,0000	470	472	479
478	1256	1257	0,0000	0	0	479

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
479	5	1256	0,0000	477	478	484
480	995	1255	0,0000	0	0	721
481	31	1254	0,0000	367	419	728
482	1251	1253	0,0000	0	0	484
483	55	1252	0,0000	0	26	799
484	5	1251	0,0000	479	482	487
485	1248	1250	0,0000	0	0	487
486	1073	1249	0,0000	0	0	650
487	5	1248	0,0000	484	485	492
488	199	1247	0,0000	355	475	1002
489	1232	1246	0,0000	0	0	501
490	1235	1245	0,0000	0	0	498
491	1242	1244	0,0000	0	0	492
492	5	1242	0,0000	487	491	506
493	449	1240	0,0000	0	359	494
494	449	1239	0,0000	493	0	1629
495	1236	1238	0,0000	0	0	497
496	1138	1237	0,0000	0	0	590
497	97	1236	0,0000	448	495	681
498	60	1235	0,0000	426	490	704
499	1210	1234	0,0000	0	0	522
500	76	1233	0,0000	364	431	811
501	14	1232	0,0000	377	489	652
502	1226	1230	0,0000	0	0	506
503	1172	1229	0,0000	0	0	557
504	1223	1228	0,0000	0	0	509
505	1175	1227	0,0000	0	0	554
506	5	1226	0,0000	492	502	510
507	1038	1225	0,0000	0	0	681
508	1222	1224	0,0000	0	0	510
509	164	1223	0,0000	0	504	1626
510	5	1222	0,0000	506	508	515
511	1217	1221	0,0000	0	0	515
512	98	1220	0,0000	408	411	993
513	913	1219	0,0000	0	0	798
514	1198	1218	0,0000	0	0	533
515	5	1217	0,0000	510	511	517
516	1215	1216	0,0000	0	0	517
517	5	1215	0,0000	515	516	519
518	1213	1214	0,0000	0	0	519
519	5	1213	0,0000	517	518	524
520	344	1212	0,0000	381	463	573
521	1207	1211	0,0000	0	0	524
522	82	1210	0,0000	476	499	567
523	1186	1209	0,0000	0	0	545
524	5	1207	0,0000	519	521	539
525	1129	1206	0,0000	0	0	598
526	1194	1205	0,0000	0	0	537
527	1162	1204	0,0000	0	0	567
528	28	1203	0,0000	0	311	1136
529	955	1202	0,0000	0	0	759
530	481	1201	0,0000	0	0	1189
531	1153	1200	0,0000	0	0	576
532	738	1199	0,0000	407	0	1630
533	22	1198	0,0000	468	514	601
534	1192	1197	0,0000	0	0	539
535	9	1196	0,0000	0	116	1113
536	335	1195	0,0000	0	0	1575
537	45	1194	0,0000	0	526	1219
538	397	1193	0,0000	0	274	1567
539	5	1192	0,0000	524	534	541
540	1190	1191	0,0000	0	0	541
541	5	1190	0,0000	539	540	547
542	1180	1189	0,0000	0	450	1589
543	1183	1188	0,0000	0	0	547
544	949	1187	0,0000	0	0	765
545	52	1186	0,0000	456	523	643
546	1156	1185	0,0000	0	0	573
547	5	1183	0,0000	541	543	566
548	938	1182	0,0000	0	0	775
549	27	1181	0,0000	432	447	639
550	61	1179	0,0000	187	282	791
551	381	1178	0,0000	0	35	942
552	1158	1177	0,0000	0	0	571
553	87	1176	0,0000	0	173	1175
554	140	1175	0,0000	31	505	725
555	44	1174	0,0000	0	373	808
556	96	1173	0,0000	137	465	730
557	6	1172	0,0000	386	503	580
558	898	1171	0,0000	0	101	746

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
559	1114	1170	0,0000	0	0	611
560	1163	1169	0,0000	0	0	566
561	1012	1168	0,0000	0	0	704
562	307	1167	0,0000	0	362	1092
563	1149	1166	0,0000	0	0	580
564	138	1165	0,0000	442	444	818
565	1112	1164	0,0000	0	0	613
566	5	1163	0,0000	547	560	575
567	82	1162	0,0000	522	527	621
568	920	1161	0,0000	0	0	791
569	252	1160	0,0000	0	429	1624
570	315	1159	0,0000	0	51	654
571	422	1158	0,0000	449	552	895
572	586	1157	0,0000	0	0	1092
573	344	1156	0,0000	520	546	751
574	1154	1155	0,0000	0	0	575
575	5	1154	0,0000	566	574	578
576	32	1153	0,0000	435	531	1096
577	1151	1152	0,0000	0	0	578
578	5	1151	0,0000	575	577	581
579	1148	1150	0,0000	0	0	581
580	6	1149	0,0000	557	563	670
581	5	1148	0,0000	578	579	586
582	1142	1147	0,0000	0	0	586
583	988	1145	0,0000	0	0	728
584	772	1144	0,0000	0	0	928
585	1051	1143	0,0000	0	0	670
586	5	1142	0,0000	581	582	592
587	682	1141	0,0000	0	60	1623
588	1135	1140	0,0000	0	0	592
589	1125	1139	0,0000	0	0	601
590	47	1138	0,0000	452	496	1210
591	13	1136	0,0000	328	453	659
592	5	1135	0,0000	586	588	600
593	929	1134	0,0000	0	0	783
594	1126	1133	0,0000	0	0	600
595	873	1132	0,0000	0	0	835
596	497	1131	0,0000	0	91	1593
597	358	1130	0,0000	0	363	1028
598	268	1129	0,0000	0	525	1073
599	1080	1128	0,0000	0	0	643
600	5	1126	0,0000	592	594	605
601	22	1125	0,0000	533	589	661
602	1120	1124	0,0000	0	0	605
603	813	1123	0,0000	0	0	891
604	392	1122	0,0000	0	473	1590
605	5	1120	0,0000	600	602	622
606	912	1119	0,0000	0	0	799
607	25	1118	0,0000	0	353	1322
608	1104	1117	0,0000	0	0	621
609	1085	1116	0,0000	0	0	639
610	684	1115	0,0000	0	0	1006
611	527	1114	0,0000	165	559	1596
612	963	1113	0,0000	0	0	751
613	159	1112	0,0000	348	565	738
614	724	1111	0,0000	0	0	970
615	1103	1110	0,0000	0	0	622
616	1064	1109	0,0000	0	0	659
617	998	1108	0,0000	0	0	718
618	196	1107	0,0000	0	0	1548
619	582	1106	0,0000	0	0	1096
620	1071	1105	0,0000	0	0	652
621	82	1104	0,0000	567	608	717
622	5	1103	0,0000	605	615	627
623	137	1101	0,0000	0	0	1573
624	740	1100	0,0000	0	0	956
625	1076	1099	0,0000	0	0	647
626	1097	1098	0,0000	0	0	627
627	5	1097	0,0000	622	626	632
628	271	1096	0,0000	467	469	879
629	39	1095	0,0000	433	0	1644
630	891	1094	0,0000	0	0	818
631	1092	1093	0,0000	0	0	632
632	5	1092	0,0000	627	631	635
633	902	1091	0,0000	0	0	808
634	1089	1090	0,0000	0	0	635
635	5	1089	0,0000	632	634	645
636	180	1088	0,0000	0	466	822
637	809	1087	0,0000	0	0	895
638	925	1086	0,0000	0	0	786

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
639	27	1085	0,0000	549	609	701
640	1054	1084	0,0000	0	0	668
641	803	1083	0,0000	0	0	900
642	688	1082	0,0000	0	0	1002
643	52	1080	0,0000	545	599	691
644	1078	1079	0,0000	0	0	645
645	5	1078	0,0000	635	644	653
646	1035	1077	0,0000	0	0	684
647	596	1076	0,0000	176	625	924
648	1070	1075	0,0000	0	0	653
649	352	1074	0,0000	0	354	1087
650	40	1073	0,0000	250	486	883
651	156	1072	0,0000	0	0	1546
652	14	1071	0,0000	501	620	750
653	5	1070	0,0000	645	648	664
654	315	1069	0,0000	570	0	1627
655	1027	1068	0,0000	0	0	691
656	1062	1067	0,0000	0	0	661
657	1059	1066	0,0000	0	0	664
658	823	1065	0,0000	0	0	884
659	13	1064	0,0000	591	616	994
660	145	1063	0,0000	0	161	1099
661	22	1062	0,0000	601	656	676
662	226	1061	0,0000	29	396	1114
663	1043	1060	0,0000	0	0	676
664	5	1059	0,0000	653	657	666
665	1056	1057	0,0000	0	0	666
666	5	1056	0,0000	664	665	672
667	689	1055	0,0000	0	40	974
668	1	1054	0,0000	394	640	1444
669	1048	1052	0,0000	0	0	672
670	6	1051	0,0000	580	585	737
671	979	1049	0,0000	0	0	737
672	5	1048	0,0000	666	669	675
673	283	1046	0,0000	191	233	860
674	1044	1045	0,0000	0	0	675
675	5	1044	0,0000	672	674	687
676	22	1043	0,0000	661	663	756
677	1031	1042	0,0000	0	0	687
678	51	1041	0,0000	0	90	1192
679	977	1040	0,0000	0	0	738
680	991	1039	0,0000	0	0	725
681	97	1038	0,0000	497	507	713
682	105	1037	0,0000	269	317	885
683	132	1036	0,0000	139	397	1043
684	86	1035	0,0000	0	646	1072
685	1016	1034	0,0000	0	0	701
686	824	1032	0,0000	0	0	883
687	5	1031	0,0000	675	677	692
688	964	1030	0,0000	0	0	750
689	1026	1029	0,0000	0	0	692
690	148	1028	0,0000	245	0	1674
691	52	1027	0,0000	643	655	806
692	5	1026	0,0000	687	689	694
693	1023	1024	0,0000	0	0	694
694	5	1023	0,0000	692	693	696
695	1021	1022	0,0000	0	0	696
696	5	1021	0,0000	694	695	698
697	1019	1020	0,0000	0	0	698
698	5	1019	0,0000	696	697	702
699	1015	1018	0,0000	0	0	702
700	48	1017	0,0000	0	237	1253
701	27	1016	0,0000	639	685	840
702	5	1015	0,0000	698	699	716
703	958	1014	0,0000	0	0	756
704	60	1012	0,0000	498	561	810
705	1000	1011	0,0000	0	0	716
706	102	1010	0,0000	0	430	723
707	1003	1009	0,0000	0	0	713
708	986	1008	0,0000	0	0	730
709	720	1007	0,0000	0	0	974
710	81	1006	0,0000	0	157	998
711	900	1005	0,0000	0	0	810
712	771	1004	0,0000	0	0	929
713	97	1003	0,0000	681	707	886
714	999	1002	0,0000	0	0	717
715	615	1001	0,0000	0	0	1549
716	5	1000	0,0000	702	705	722
717	82	999	0,0000	621	714	1180
718	89	998	0,0000	458	617	1415
719	994	997	0,0000	0	0	722

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
720	858	996	0,0000	0	0	849
721	85	995	0,0000	451	480	989
722	5	994	0,0000	716	719	729
723	102	993	0,0000	706	0	1616
724	849	992	0,0000	0	0	858
725	140	991	0,0000	554	680	1340
726	987	990	0,0000	0	0	729
727	699	989	0,0000	0	0	993
728	31	988	0,0000	481	583	1205
729	5	987	0,0000	722	726	733
730	96	986	0,0000	556	708	740
731	405	985	0,0000	0	0	1253
732	983	984	0,0000	0	0	733
733	5	983	0,0000	729	732	755
734	868	982	0,0000	0	0	840
735	975	981	0,0000	0	0	740
736	846	980	0,0000	0	0	861
737	6	979	0,0000	670	671	868
738	159	977	0,0000	613	679	932
739	78	976	0,0000	349	474	1086
740	96	975	0,0000	730	735	1011
741	774	974	0,0000	0	0	926
742	959	973	0,0000	0	0	755
743	197	972	0,0000	0	321	1599
744	859	971	0,0000	0	0	848
745	847	970	0,0000	0	0	860
746	898	969	0,0000	558	0	1623
747	24	968	0,0000	0	120	1229
748	579	967	0,0000	0	0	1099
749	542	966	0,0000	0	0	1136
750	14	964	0,0000	652	688	832
751	344	963	0,0000	573	612	968
752	165	962	0,0000	0	0	1573
753	662	961	0,0000	0	0	1547
754	703	960	0,0000	0	0	990
755	5	959	0,0000	733	742	764
756	22	958	0,0000	676	703	770
757	944	957	0,0000	0	0	770
758	506	956	0,0000	0	0	1167
759	91	955	0,0000	358	529	1207
760	950	954	0,0000	0	0	764
761	762	953	0,0000	0	28	1557
762	131	952	0,0000	115	324	1223
763	808	951	0,0000	0	0	896
764	5	950	0,0000	755	760	771
765	221	949	0,0000	0	544	801
766	909	948	0,0000	0	0	801
767	876	947	0,0000	0	0	832
768	942	946	0,0000	0	0	771
769	466	945	0,0000	0	455	1117
770	22	944	0,0000	756	757	813
771	5	942	0,0000	764	768	778
772	275	941	0,0000	0	27	955
773	839	940	0,0000	0	0	868
774	896	939	0,0000	0	0	813
775	232	938	0,0000	0	548	880
776	935	937	0,0000	0	0	778
777	802	936	0,0000	0	0	901
778	5	935	0,0000	771	776	781
779	931	933	0,0000	0	0	781
780	924	932	0,0000	0	0	787
781	5	931	0,0000	778	779	793
782	192	930	0,0000	0	99	1018
783	476	929	0,0000	0	593	1603
784	918	927	0,0000	0	0	793
785	491	926	0,0000	0	0	1180
786	401	925	0,0000	41	638	1120
787	183	924	0,0000	240	780	869
788	904	923	0,0000	0	0	806
789	290	922	0,0000	0	388	1283
790	495	921	0,0000	0	273	816
791	61	920	0,0000	550	568	1145
792	725	919	0,0000	0	0	969
793	5	918	0,0000	781	784	795
794	916	917	0,0000	0	0	795
795	5	916	0,0000	793	794	819
796	565	915	0,0000	0	0	1113
797	726	914	0,0000	0	0	968
798	258	913	0,0000	0	513	1256
799	55	912	0,0000	483	606	1635
800	899	910	0,0000	0	0	811

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
801	221	909	0,0000	765	766	1386
802	890	908	0,0000	0	0	819
803	35	907	0,0000	121	148	1446
804	844	906	0,0000	0	0	863
805	671	905	0,0000	0	0	1018
806	52	904	0,0000	691	788	890
807	593	903	0,0000	0	0	1086
808	44	902	0,0000	555	633	872
809	389	901	0,0000	0	0	1265
810	60	900	0,0000	704	711	873
811	76	899	0,0000	500	800	1344
812	723	897	0,0000	0	0	971
813	22	896	0,0000	770	774	831
814	739	895	0,0000	0	0	1569
815	644	894	0,0000	0	0	1043
816	495	893	0,0000	790	0	1612
817	777	892	0,0000	0	0	924
818	138	891	0,0000	564	630	1391
819	5	890	0,0000	795	802	825
820	820	889	0,0000	0	0	886
821	95	888	0,0000	0	0	1552
822	180	887	0,0000	636	0	1646
823	884	886	0,0000	0	0	825
824	877	885	0,0000	0	0	831
825	5	884	0,0000	819	823	829
826	880	883	0,0000	0	0	829
827	814	882	0,0000	0	0	890
828	835	881	0,0000	0	0	872
829	5	880	0,0000	825	826	833
830	875	878	0,0000	0	0	833
831	22	877	0,0000	813	824	865
832	14	876	0,0000	750	767	953
833	5	875	0,0000	829	830	838
834	686	874	0,0000	0	0	1004
835	175	873	0,0000	342	595	967
836	870	872	0,0000	0	0	838
837	834	871	0,0000	0	0	873
838	5	870	0,0000	833	836	852
839	855	869	0,0000	0	0	852
840	27	868	0,0000	701	734	889
841	455	866	0,0000	0	0	1210
842	775	865	0,0000	0	0	1571
843	149	864	0,0000	0	0	1564
844	767	863	0,0000	0	0	932
845	838	862	0,0000	0	0	869
846	828	861	0,0000	0	0	879
847	822	860	0,0000	0	0	885
848	399	859	0,0000	113	744	1644
849	144	858	0,0000	252	720	1650
850	605	857	0,0000	170	0	1626
851	213	856	0,0000	0	0	1391
852	5	855	0,0000	838	839	864
853	704	854	0,0000	0	0	989
854	531	853	0,0000	0	0	1145
855	16	852	0,0000	0	34	1291
856	842	851	0,0000	0	0	865
857	294	850	0,0000	0	0	1553
858	339	849	0,0000	0	724	1591
859	843	848	0,0000	0	0	864
860	283	847	0,0000	673	745	1159
861	383	846	0,0000	126	736	1245
862	825	845	0,0000	0	0	882
863	461	844	0,0000	0	804	1583
864	5	843	0,0000	852	859	870
865	22	842	0,0000	831	856	892
866	837	841	0,0000	0	0	870
867	727	840	0,0000	0	0	967
868	6	839	0,0000	737	773	952
869	183	838	0,0000	787	845	1153
870	5	837	0,0000	864	866	887
871	785	836	0,0000	0	0	917
872	44	835	0,0000	808	828	1082
873	60	834	0,0000	810	837	988
874	812	833	0,0000	0	0	892
875	819	832	0,0000	0	0	887
876	698	831	0,0000	0	0	994
877	827	830	0,0000	0	0	880
878	815	829	0,0000	0	0	889
879	271	828	0,0000	628	846	1098
880	232	827	0,0000	775	877	1337
881	292	826	0,0000	0	0	1344

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
882	764	825	0,0000	0	862	1587
883	40	824	0,0000	650	686	1037
884	2	823	0,0000	0	658	1443
885	105	822	0,0000	682	847	1053
886	97	820	0,0000	713	820	1163
887	5	819	0,0000	870	875	903
888	770	816	0,0000	288	424	1637
889	27	815	0,0000	840	878	951
890	52	814	0,0000	806	827	930
891	15	813	0,0000	409	603	1204
892	22	812	0,0000	865	874	907
893	755	811	0,0000	0	0	942
894	646	810	0,0000	0	0	1041
895	422	809	0,0000	571	637	1172
896	166	808	0,0000	56	763	1081
897	800	807	0,0000	0	0	903
898	450	806	0,0000	235	383	927
899	795	805	0,0000	0	0	907
900	746	803	0,0000	371	641	1679
901	238	802	0,0000	146	777	1367
902	769	801	0,0000	0	0	930
903	5	800	0,0000	887	897	909
904	744	799	0,0000	0	0	952
905	173	798	0,0000	0	63	1328
906	793	797	0,0000	0	0	909
907	22	795	0,0000	892	899	1014
908	72	794	0,0000	0	278	1577
909	5	793	0,0000	903	906	911
910	791	792	0,0000	0	0	911
911	5	791	0,0000	909	910	913
912	789	790	0,0000	0	0	913
913	5	789	0,0000	911	912	915
914	787	788	0,0000	0	0	915
915	5	787	0,0000	913	914	921
916	602	786	0,0000	0	0	1081
917	228	785	0,0000	0	871	1173
918	781	784	0,0000	0	0	921
919	660	783	0,0000	0	0	1028
920	379	782	0,0000	0	0	1273
921	5	781	0,0000	915	918	925
922	773	780	0,0000	0	0	927
923	776	779	0,0000	0	0	925
924	596	777	0,0000	647	817	991
925	5	776	0,0000	921	923	940
926	328	774	0,0000	0	741	1267
927	450	773	0,0000	898	922	1182
928	363	772	0,0000	441	584	1143
929	260	771	0,0000	0	712	1128
930	52	769	0,0000	890	902	934
931	757	768	0,0000	0	0	940
932	159	767	0,0000	738	844	1187
933	765	766	0,0000	0	0	934
934	52	765	0,0000	930	933	1035
935	462	763	0,0000	0	0	1204
936	741	761	0,0000	0	0	955
937	533	760	0,0000	0	0	1143
938	224	759	0,0000	0	0	1593
939	489	758	0,0000	0	0	1182
940	5	757	0,0000	925	931	946
941	416	756	0,0000	0	0	1243
942	381	755	0,0000	551	893	1641
943	705	754	0,0000	0	0	988
944	751	753	0,0000	0	0	946
945	282	752	0,0000	0	0	1351
946	5	751	0,0000	940	944	957
947	743	750	0,0000	0	0	953
948	737	749	0,0000	0	0	957
949	745	748	0,0000	0	0	951
950	624	747	0,0000	0	0	1063
951	27	745	0,0000	889	949	1016
952	6	744	0,0000	868	904	1188
953	14	743	0,0000	832	947	995
954	693	742	0,0000	0	0	998
955	275	741	0,0000	772	936	1278
956	540	740	0,0000	375	624	1681
957	5	737	0,0000	946	948	959
958	735	736	0,0000	0	0	959
959	5	735	0,0000	957	958	963
960	501	734	0,0000	0	0	1172
961	731	733	0,0000	0	0	963
962	580	732	0,0000	0	0	1098

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
963	5	731	0,0000	959	961	965
964	729	730	0,0000	0	0	965
965	5	729	0,0000	963	964	979
966	666	728	0,0000	0	0	1023
967	175	727	0,0000	835	867	1022
968	344	726	0,0000	751	797	1062
969	114	725	0,0000	0	792	1600
970	65	724	0,0000	0	614	1431
971	246	723	0,0000	330	812	1091
972	550	722	0,0000	0	0	1128
973	650	721	0,0000	0	0	1037
974	689	720	0,0000	667	709	1628
975	634	719	0,0000	0	0	1053
976	696	718	0,0000	0	0	995
977	128	717	0,0000	0	0	1443
978	714	716	0,0000	0	0	979
979	5	714	0,0000	965	978	986
980	675	713	0,0000	0	0	1014
981	702	712	0,0000	0	0	991
982	707	711	0,0000	0	0	986
983	678	710	0,0000	0	0	1011
984	673	709	0,0000	0	0	1016
985	652	708	0,0000	0	0	1035
986	5	707	0,0000	979	982	997
987	413	706	0,0000	0	0	1245
988	60	705	0,0000	873	943	1049
989	85	704	0,0000	721	853	1149
990	254	703	0,0000	410	754	1720
991	596	702	0,0000	924	981	1055
992	694	701	0,0000	0	0	997
993	98	699	0,0000	512	727	1077
994	13	698	0,0000	659	876	1224
995	14	696	0,0000	953	976	1061
996	607	695	0,0000	0	0	1077
997	5	694	0,0000	986	992	1003
998	81	693	0,0000	710	954	1194
999	600	692	0,0000	0	0	1082
1000	687	691	0,0000	0	0	1003
1001	269	690	0,0000	0	0	1579
1002	199	688	0,0000	488	642	1695
1003	5	687	0,0000	997	1000	1013
1004	209	686	0,0000	0	834	1594
1005	478	685	0,0000	0	0	1192
1006	309	684	0,0000	0	610	1625
1007	477	683	0,0000	0	0	1575
1008	510	681	0,0000	0	0	1163
1009	612	680	0,0000	0	0	1073
1010	170	679	0,0000	0	0	1560
1011	96	678	0,0000	740	983	1064
1012	676	677	0,0000	0	0	1013
1013	5	676	0,0000	1003	1012	1017
1014	22	675	0,0000	907	980	1042
1015	672	674	0,0000	0	0	1017
1016	27	673	0,0000	951	984	1065
1017	5	672	0,0000	1013	1015	1024
1018	192	671	0,0000	782	805	1640
1019	626	670	0,0000	0	0	1061
1020	665	669	0,0000	0	0	1024
1021	346	668	0,0000	0	0	1302
1022	175	667	0,0000	967	0	1664
1023	597	666	0,0000	368	966	1618
1024	5	665	0,0000	1017	1020	1026
1025	663	664	0,0000	0	0	1026
1026	5	663	0,0000	1024	1025	1034
1027	266	661	0,0000	0	0	1565
1028	358	660	0,0000	597	919	1074
1029	645	658	0,0000	0	0	1042
1030	522	657	0,0000	0	0	1153
1031	621	656	0,0000	0	0	1065
1032	653	655	0,0000	0	0	1034
1033	68	654	0,0000	0	82	1474
1034	5	653	0,0000	1026	1032	1044
1035	52	652	0,0000	934	985	1123
1036	643	651	0,0000	0	0	1044
1037	40	650	0,0000	883	973	1332
1038	638	649	0,0000	0	0	1049
1039	625	648	0,0000	0	0	1062
1040	498	647	0,0000	0	0	1175
1041	4	646	0,0000	406	894	1119
1042	22	645	0,0000	1014	1029	1048
1043	132	644	0,0000	683	815	1373

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
882	764	825	0,0000	0	862	1587
883	40	824	0,0000	650	686	1037
884	2	823	0,0000	0	658	1443
885	105	822	0,0000	682	847	1053
886	97	820	0,0000	713	820	1163
887	5	819	0,0000	870	875	903
888	770	816	0,0000	288	424	1637
889	27	815	0,0000	840	878	951
890	52	814	0,0000	806	827	930
891	15	813	0,0000	409	603	1204
892	22	812	0,0000	865	874	907
893	755	811	0,0000	0	0	942
894	646	810	0,0000	0	0	1041
895	422	809	0,0000	571	637	1172
896	166	808	0,0000	56	763	1081
897	800	807	0,0000	0	0	903
898	450	806	0,0000	235	383	927
899	795	805	0,0000	0	0	907
900	746	803	0,0000	371	641	1679
901	238	802	0,0000	146	777	1367
902	769	801	0,0000	0	0	930
903	5	800	0,0000	887	897	909
904	744	799	0,0000	0	0	952
905	173	798	0,0000	0	63	1328
906	793	797	0,0000	0	0	909
907	22	795	0,0000	892	899	1014
908	72	794	0,0000	0	278	1577
909	5	793	0,0000	903	906	911
910	791	792	0,0000	0	0	911
911	5	791	0,0000	909	910	913
912	789	790	0,0000	0	0	913
913	5	789	0,0000	911	912	915
914	787	788	0,0000	0	0	915
915	5	787	0,0000	913	914	921
916	602	786	0,0000	0	0	1081
917	228	785	0,0000	0	871	1173
918	781	784	0,0000	0	0	921
919	660	783	0,0000	0	0	1028
920	379	782	0,0000	0	0	1273
921	5	781	0,0000	915	918	925
922	773	780	0,0000	0	0	927
923	776	779	0,0000	0	0	925
924	596	777	0,0000	647	817	991
925	5	776	0,0000	921	923	940
926	328	774	0,0000	0	741	1267
927	450	773	0,0000	898	922	1182
928	363	772	0,0000	441	584	1143
929	260	771	0,0000	0	712	1128
930	52	769	0,0000	890	902	934
931	757	768	0,0000	0	0	940
932	159	767	0,0000	738	844	1187
933	765	766	0,0000	0	0	934
934	52	765	0,0000	930	933	1035
935	462	763	0,0000	0	0	1204
936	741	761	0,0000	0	0	955
937	533	760	0,0000	0	0	1143
938	224	759	0,0000	0	0	1593
939	489	758	0,0000	0	0	1182
940	5	757	0,0000	925	931	946
941	416	756	0,0000	0	0	1243
942	381	755	0,0000	551	893	1641
943	705	754	0,0000	0	0	988
944	751	753	0,0000	0	0	946
945	282	752	0,0000	0	0	1351
946	5	751	0,0000	940	944	957
947	743	750	0,0000	0	0	953
948	737	749	0,0000	0	0	957
949	745	748	0,0000	0	0	951
950	624	747	0,0000	0	0	1063
951	27	745	0,0000	889	949	1016
952	6	744	0,0000	868	904	1188
953	14	743	0,0000	832	947	995
954	693	742	0,0000	0	0	998
955	275	741	0,0000	772	936	1278
956	540	740	0,0000	375	624	1681
957	5	737	0,0000	946	948	959
958	735	736	0,0000	0	0	959
959	5	735	0,0000	957	958	963
960	501	734	0,0000	0	0	1172
961	731	733	0,0000	0	0	963
962	580	732	0,0000	0	0	1098

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1044	5	643	0,0000	1034	1036	1051
1045	639	642	0,0000	0	0	1048
1046	319	641	0,0000	0	0	1322
1047	636	640	0,0000	0	0	1051
1048	22	639	0,0000	1042	1045	1095
1049	60	638	0,0000	988	1038	1155
1050	482	637	0,0000	0	0	1188
1051	5	636	0,0000	1044	1047	1067
1052	317	635	0,0000	0	12	1632
1053	105	634	0,0000	885	975	1280
1054	500	633	0,0000	0	0	1173
1055	596	632	0,0000	991	0	1692
1056	592	631	0,0000	0	0	1087
1057	139	630	0,0000	0	66	1588
1058	622	629	0,0000	0	0	1064
1059	194	628	0,0000	0	0	1405
1060	619	627	0,0000	0	0	1067
1061	14	626	0,0000	995	1019	1070
1062	344	625	0,0000	968	1039	1142
1063	273	624	0,0000	283	950	1613
1064	96	622	0,0000	1011	1058	1284
1065	27	621	0,0000	1016	1031	1130
1066	441	620	0,0000	0	0	1219
1067	5	619	0,0000	1051	1060	1085
1068	616	618	0,0000	0	0	1070
1069	548	617	0,0000	0	0	1130
1070	14	616	0,0000	1061	1068	1165
1071	451	614	0,0000	0	0	1591
1072	86	613	0,0000	684	0	1621
1073	268	612	0,0000	598	1009	1346
1074	358	611	0,0000	1028	0	1651
1075	124	609	0,0000	374	420	1262
1076	520	608	0,0000	0	0	1155
1077	98	607	0,0000	993	996	1459
1078	272	606	0,0000	0	167	1592
1079	534	604	0,0000	0	0	1142
1080	559	603	0,0000	0	0	1119
1081	166	602	0,0000	896	916	1368
1082	44	600	0,0000	872	999	1432
1083	508	599	0,0000	0	0	1165
1084	594	595	0,0000	0	0	1085
1085	5	594	0,0000	1067	1084	1093
1086	78	593	0,0000	739	807	1274
1087	352	592	0,0000	649	1056	1251
1088	583	591	0,0000	0	0	1095
1089	436	590	0,0000	0	0	1224
1090	585	588	0,0000	0	0	1093
1091	246	587	0,0000	971	0	1665
1092	307	586	0,0000	562	572	1632
1093	5	585	0,0000	1085	1090	1102
1094	555	584	0,0000	0	0	1123
1095	22	583	0,0000	1048	1088	1126
1096	32	582	0,0000	576	619	1231
1097	374	581	0,0000	0	0	1278
1098	271	580	0,0000	879	962	1226
1099	145	579	0,0000	660	748	1246
1100	576	578	0,0000	0	0	1102
1101	393	577	0,0000	0	0	1262
1102	5	576	0,0000	1093	1100	1107
1103	233	575	0,0000	0	0	1561
1104	571	574	0,0000	0	0	1107
1105	429	573	0,0000	0	0	1231
1106	557	572	0,0000	0	0	1121
1107	5	571	0,0000	1102	1104	1111
1108	127	570	0,0000	0	0	1444
1109	305	569	0,0000	0	0	1332
1110	567	568	0,0000	0	0	1111
1111	5	567	0,0000	1107	1110	1115
1112	563	566	0,0000	0	0	1115
1113	9	565	0,0000	535	796	1255
1114	226	564	0,0000	662	0	1667
1115	5	563	0,0000	1111	1112	1132
1116	458	562	0,0000	0	0	1207
1117	466	561	0,0000	769	0	1700
1118	360	560	0,0000	0	0	1291
1119	4	559	0,0000	1041	1080	1160
1120	401	558	0,0000	786	0	1648
1121	324	557	0,0000	0	1106	1586
1122	516	556	0,0000	0	0	1159
1123	52	555	0,0000	1035	1094	1193
1124	552	554	0,0000	0	0	1126

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1125	483	553	0,0000	0	0	1187
1126	22	552	0,0000	1095	1124	1129
1127	549	551	0,0000	0	0	1129
1128	260	550	0,0000	929	972	1645
1129	22	549	0,0000	1126	1127	1138
1130	27	548	0,0000	1065	1069	1212
1131	546	547	0,0000	0	0	1132
1132	5	546	0,0000	1115	1131	1134
1133	544	545	0,0000	0	0	1134
1134	5	544	0,0000	1132	1133	1141
1135	539	543	0,0000	0	0	1138
1136	28	542	0,0000	528	749	1634
1137	249	541	0,0000	0	222	1619
1138	22	539	0,0000	1129	1135	1158
1139	536	538	0,0000	0	0	1141
1140	517	537	0,0000	0	0	1158
1141	5	536	0,0000	1134	1139	1146
1142	344	534	0,0000	1062	1079	1176
1143	363	533	0,0000	928	937	1181
1144	530	532	0,0000	0	0	1146
1145	61	531	0,0000	791	854	1282
1146	5	530	0,0000	1141	1144	1156
1147	526	529	0,0000	0	0	1149
1148	519	528	0,0000	0	0	1156
1149	85	526	0,0000	989	1147	1377
1150	378	525	0,0000	0	0	1274
1151	514	524	0,0000	0	0	1160
1152	297	523	0,0000	0	0	1340
1153	183	522	0,0000	869	1030	1277
1154	496	521	0,0000	0	0	1176
1155	60	520	0,0000	1049	1076	1235
1156	5	519	0,0000	1146	1148	1161
1157	513	518	0,0000	0	0	1161
1158	22	517	0,0000	1138	1140	1222
1159	283	516	0,0000	860	1122	1333
1160	4	514	0,0000	1119	1151	1299
1161	5	513	0,0000	1156	1157	1169
1162	212	511	0,0000	0	0	1562
1163	97	510	0,0000	886	1008	1411
1164	504	509	0,0000	0	0	1169
1165	14	508	0,0000	1070	1083	1298
1166	474	507	0,0000	0	0	1194
1167	162	506	0,0000	268	758	1211
1168	438	505	0,0000	0	0	1222
1169	5	504	0,0000	1161	1164	1177
1170	350	503	0,0000	0	0	1298
1171	494	502	0,0000	0	0	1177
1172	422	501	0,0000	895	960	1659
1173	228	500	0,0000	917	1054	1196
1174	403	499	0,0000	0	0	1255
1175	87	498	0,0000	553	1040	1656
1176	344	496	0,0000	1142	1154	1221
1177	5	494	0,0000	1169	1171	1179
1178	492	493	0,0000	0	0	1179
1179	5	492	0,0000	1177	1178	1185
1180	82	491	0,0000	717	785	1360
1181	363	490	0,0000	1143	0	1670
1182	450	489	0,0000	927	939	1699
1183	485	488	0,0000	0	0	1185
1184	475	486	0,0000	0	0	1193
1185	5	485	0,0000	1179	1183	1198
1186	289	484	0,0000	0	0	1346
1187	159	483	0,0000	932	1125	1341
1188	6	482	0,0000	952	1050	1309
1189	415	481	0,0000	457	530	1636
1190	460	480	0,0000	0	0	1205
1191	469	479	0,0000	0	0	1198
1192	51	478	0,0000	678	1005	1385
1193	52	475	0,0000	1123	1184	1260
1194	81	474	0,0000	998	1166	1347
1195	448	473	0,0000	0	0	1212
1196	228	472	0,0000	1173	0	1584
1197	454	471	0,0000	0	0	1211
1198	5	469	0,0000	1185	1191	1202
1199	464	468	0,0000	0	0	1202
1200	214	467	0,0000	376	0	1601
1201	84	465	0,0000	0	149	1611
1202	5	464	0,0000	1198	1199	1208
1203	375	463	0,0000	0	0	1277
1204	15	462	0,0000	891	935	1493
1205	31	460	0,0000	728	1190	1353

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1206	457	459	0,0000	0	0	1208
1207	91	458	0,0000	759	1116	1428
1208	5	457	0,0000	1202	1206	1216
1209	349	456	0,0000	0	0	1299
1210	47	455	0,0000	590	841	1452
1211	162	454	0,0000	1167	1197	1715
1212	27	448	0,0000	1130	1195	1376
1213	444	447	0,0000	0	0	1216
1214	109	446	0,0000	0	0	1459
1215	300	445	0,0000	0	0	1337
1216	5	444	0,0000	1208	1213	1218
1217	442	443	0,0000	0	0	1218
1218	5	442	0,0000	1216	1217	1225
1219	45	441	0,0000	537	1066	1396
1220	435	440	0,0000	0	0	1225
1221	344	439	0,0000	1176	0	1737
1222	22	438	0,0000	1158	1168	1250
1223	131	437	0,0000	762	0	1658
1224	13	436	0,0000	994	1089	1706
1225	5	435	0,0000	1218	1220	1230
1226	271	434	0,0000	1098	0	1698
1227	49	433	0,0000	0	264	1607
1228	430	432	0,0000	0	0	1230
1229	24	431	0,0000	747	0	1610
1230	5	430	0,0000	1225	1228	1236
1231	32	429	0,0000	1096	1105	1329
1232	424	428	0,0000	0	0	1236
1233	425	427	0,0000	0	0	1235
1234	412	426	0,0000	0	0	1246
1235	60	425	0,0000	1155	1233	1241
1236	5	424	0,0000	1230	1232	1248
1237	395	423	0,0000	0	0	1260
1238	103	421	0,0000	0	0	1574
1239	410	420	0,0000	0	0	1248
1240	418	419	0,0000	0	0	1241
1241	60	418	0,0000	1235	1240	1271
1242	408	417	0,0000	0	0	1250
1243	143	416	0,0000	401	941	1606
1244	172	414	0,0000	0	217	1383
1245	383	413	0,0000	861	987	1652
1246	145	412	0,0000	1099	1234	1657
1247	368	411	0,0000	0	0	1284
1248	5	410	0,0000	1236	1239	1257
1249	400	409	0,0000	0	0	1257
1250	22	408	0,0000	1222	1242	1288
1251	352	407	0,0000	1087	0	1647
1252	187	406	0,0000	0	0	1411
1253	48	405	0,0000	700	731	1412
1254	310	404	0,0000	0	0	1329
1255	9	403	0,0000	1113	1174	1667
1256	258	402	0,0000	798	0	1617
1257	5	400	0,0000	1248	1249	1259
1258	396	398	0,0000	0	0	1259
1259	5	396	0,0000	1257	1258	1269
1260	52	395	0,0000	1193	1237	1372
1261	382	394	0,0000	0	0	1271
1262	124	393	0,0000	1075	1101	1335
1263	385	391	0,0000	0	0	1269
1264	337	390	0,0000	0	0	1309
1265	321	389	0,0000	0	809	1595
1266	206	388	0,0000	0	0	1396
1267	328	387	0,0000	926	0	1624
1268	364	386	0,0000	0	0	1288
1269	5	385	0,0000	1259	1263	1272
1270	380	384	0,0000	0	0	1272
1271	60	382	0,0000	1241	1261	1313
1272	5	380	0,0000	1269	1270	1276
1273	244	379	0,0000	37	920	1642
1274	78	378	0,0000	1086	1150	1693
1275	376	377	0,0000	0	0	1276
1276	5	376	0,0000	1272	1275	1287
1277	183	375	0,0000	1153	1203	1285
1278	275	374	0,0000	955	1097	1661
1279	365	373	0,0000	0	0	1287
1280	105	372	0,0000	1053	0	1694
1281	313	371	0,0000	0	0	1326
1282	61	370	0,0000	1145	0	1697
1283	290	369	0,0000	789	0	1643
1284	96	368	0,0000	1064	1247	1366
1285	183	367	0,0000	1277	0	1708
1286	280	366	0,0000	0	0	1353

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1287	5	365	0,0000	1276	1279	1290
1288	22	364	0,0000	1250	1268	1306
1289	361	362	0,0000	0	0	1290
1290	5	361	0,0000	1287	1289	1294
1291	16	360	0,0000	855	1118	1467
1292	331	359	0,0000	0	0	1313
1293	356	357	0,0000	0	0	1294
1294	5	356	0,0000	1290	1293	1297
1295	351	355	0,0000	0	0	1297
1296	341	354	0,0000	0	0	1306
1297	5	351	0,0000	1294	1295	1301
1298	14	350	0,0000	1165	1170	1357
1299	4	349	0,0000	1160	1209	1417
1300	347	348	0,0000	0	0	1301
1301	5	347	0,0000	1297	1300	1310
1302	219	346	0,0000	0	1021	1608
1303	182	345	0,0000	0	0	1415
1304	71	343	0,0000	361	0	1585
1305	151	342	0,0000	0	0	1432
1306	22	341	0,0000	1288	1296	1363
1307	236	340	0,0000	0	0	1377
1308	334	338	0,0000	0	0	1310
1309	6	337	0,0000	1188	1264	1462
1310	5	334	0,0000	1301	1308	1312
1311	332	333	0,0000	0	0	1312
1312	5	332	0,0000	1310	1311	1315
1313	60	331	0,0000	1271	1292	1330
1314	329	330	0,0000	0	0	1315
1315	5	329	0,0000	1312	1314	1321
1316	308	327	0,0000	0	0	1330
1317	296	326	0,0000	0	0	1341
1318	288	325	0,0000	0	0	1347
1319	320	323	0,0000	0	0	1321
1320	237	322	0,0000	0	0	1376
1321	5	320	0,0000	1315	1319	1331
1322	25	319	0,0000	607	1046	1454
1323	306	318	0,0000	0	0	1331
1324	267	316	0,0000	0	0	1357
1325	253	314	0,0000	0	0	1366
1326	3	313	0,0000	404	1281	1655
1327	302	312	0,0000	0	0	1335
1328	173	311	0,0000	905	0	1633
1329	32	310	0,0000	1231	1254	1699
1330	60	308	0,0000	1313	1316	1416
1331	5	306	0,0000	1321	1323	1338
1332	40	305	0,0000	1037	1109	1370
1333	283	304	0,0000	1159	0	1691
1334	299	303	0,0000	0	0	1338
1335	124	302	0,0000	1262	1327	1676
1336	262	301	0,0000	0	0	1360
1337	232	300	0,0000	880	1215	1666
1338	5	299	0,0000	1331	1334	1343
1339	242	298	0,0000	0	0	1372
1340	140	297	0,0000	725	1152	1390
1341	159	296	0,0000	1187	1317	1413
1342	293	295	0,0000	0	0	1343
1343	5	293	0,0000	1338	1342	1350
1344	76	292	0,0000	811	881	1683
1345	285	291	0,0000	0	0	1350
1346	268	289	0,0000	1073	1186	1665
1347	81	288	0,0000	1194	1318	1395
1348	43	287	0,0000	0	395	1455
1349	77	286	0,0000	0	0	1474
1350	5	285	0,0000	1343	1345	1354
1351	178	282	0,0000	0	945	1608
1352	278	281	0,0000	0	0	1354
1353	31	280	0,0000	1205	1286	1471
1354	5	278	0,0000	1350	1352	1358
1355	264	277	0,0000	0	0	1358
1356	125	276	0,0000	0	0	1446
1357	14	267	0,0000	1298	1324	1393
1358	5	264	0,0000	1354	1355	1361
1359	261	263	0,0000	0	0	1361
1360	82	262	0,0000	1180	1336	1714
1361	5	261	0,0000	1358	1359	1375
1362	257	259	0,0000	0	0	1363
1363	22	257	0,0000	1306	1362	1374
1364	179	256	0,0000	0	0	1417
1365	185	255	0,0000	0	0	1413
1366	96	253	0,0000	1284	1325	1718
1367	238	250	0,0000	901	0	1646

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1368	166	248	0,0000	1081	0	1671
1369	239	247	0,0000	0	0	1375
1370	40	245	0,0000	1332	0	1684
1371	240	243	0,0000	0	0	1374
1372	52	242	0,0000	1260	1339	1488
1373	132	241	0,0000	1043	0	1690
1374	22	240	0,0000	1363	1371	1399
1375	5	239	0,0000	1361	1369	1384
1376	27	237	0,0000	1212	1320	1422
1377	85	236	0,0000	1149	1307	1463
1378	203	235	0,0000	0	0	1399
1379	210	234	0,0000	0	0	1393
1380	181	231	0,0000	0	0	1416
1381	118	230	0,0000	0	0	1452
1382	225	229	0,0000	0	0	1384
1383	172	227	0,0000	1244	0	1625
1384	5	225	0,0000	1375	1382	1394
1385	51	223	0,0000	1192	0	1614
1386	221	222	0,0000	801	0	1654
1387	208	220	0,0000	0	0	1394
1388	113	218	0,0000	0	0	1455
1389	186	217	0,0000	0	0	1412
1390	140	216	0,0000	1340	0	1696
1391	138	213	0,0000	818	851	1693
1392	94	211	0,0000	0	0	1467
1393	14	210	0,0000	1357	1379	1414
1394	5	208	0,0000	1384	1387	1398
1395	81	207	0,0000	1347	0	1688
1396	45	206	0,0000	1219	1266	1486
1397	204	205	0,0000	0	0	1398
1398	5	204	0,0000	1394	1397	1401
1399	22	203	0,0000	1374	1378	1427
1400	201	202	0,0000	0	0	1401
1401	5	201	0,0000	1398	1400	1406
1402	169	200	0,0000	0	0	1422
1403	184	198	0,0000	0	0	1414
1404	193	195	0,0000	0	0	1406
1405	8	194	0,0000	271	1059	1639
1406	5	193	0,0000	1401	1404	1409
1407	152	191	0,0000	0	0	1431
1408	189	190	0,0000	0	0	1409
1409	5	189	0,0000	1406	1408	1423
1410	167	188	0,0000	0	0	1423
1411	97	187	0,0000	1163	1252	1458
1412	48	186	0,0000	1253	1389	1418
1413	159	185	0,0000	1341	1365	1419
1414	14	184	0,0000	1393	1403	1479
1415	89	182	0,0000	718	1303	1660
1416	60	181	0,0000	1330	1380	1436
1417	4	179	0,0000	1299	1364	1464
1418	48	177	0,0000	1412	0	1679
1419	159	176	0,0000	1413	0	1718
1420	56	174	0,0000	0	0	1486
1421	83	171	0,0000	0	0	1471
1422	27	169	0,0000	1376	1402	1435
1423	5	167	0,0000	1409	1410	1430
1424	153	163	0,0000	0	0	1430
1425	157	160	0,0000	0	0	1427
1426	142	158	0,0000	0	0	1436
1427	22	157	0,0000	1399	1425	1438
1428	91	155	0,0000	1207	0	1674
1429	110	154	0,0000	0	0	1458
1430	5	153	0,0000	1423	1424	1448
1431	65	152	0,0000	970	1407	1620
1432	44	151	0,0000	1082	1305	1473
1433	135	150	0,0000	0	0	1438
1434	101	147	0,0000	0	0	1464
1435	27	146	0,0000	1422	0	1731
1436	60	142	0,0000	1416	1426	1439
1437	134	141	0,0000	0	0	1439
1438	22	135	0,0000	1427	1433	1441
1439	60	134	0,0000	1436	1437	1466
1440	130	133	0,0000	0	0	1441
1441	22	130	0,0000	1438	1440	1450
1442	122	129	0,0000	0	0	1448
1443	2	128	0,0000	884	977	1597
1444	1	127	0,0000	668	1108	1675
1445	120	126	0,0000	0	0	1450
1446	35	125	0,0000	803	1356	1615
1447	79	123	0,0000	0	0	1473
1448	5	122	0,0000	1430	1442	1451

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1449	119	121	0,0000	0	0	1451
1450	22	120	0,0000	1441	1445	1470
1451	5	119	0,0000	1448	1449	1457
1452	47	118	0,0000	1210	1381	1696
1453	111	117	0,0000	0	0	1457
1454	25	116	0,0000	1322	0	1666
1455	43	113	0,0000	1348	1388	1643
1456	88	112	0,0000	0	0	1470
1457	5	111	0,0000	1451	1453	1468
1458	97	110	0,0000	1411	1429	1726
1459	98	109	0,0000	1077	1214	1465
1460	92	108	0,0000	0	0	1468
1461	99	107	0,0000	0	0	1466
1462	6	106	0,0000	1309	0	1735
1463	85	104	0,0000	1377	0	1712
1464	4	101	0,0000	1417	1434	1500
1465	98	100	0,0000	1459	0	1717
1466	60	99	0,0000	1439	1461	1743
1467	16	94	0,0000	1291	1392	1673
1468	5	92	0,0000	1457	1460	1475
1469	75	90	0,0000	0	0	1475
1470	22	88	0,0000	1450	1456	1499
1471	31	83	0,0000	1353	1421	1492
1472	67	80	0,0000	0	0	1479
1473	44	79	0,0000	1432	1447	1711
1474	68	77	0,0000	1033	1349	1650
1475	5	75	0,0000	1468	1469	1477
1476	73	74	0,0000	0	0	1477
1477	5	73	0,0000	1475	1476	1482
1478	63	69	0,0000	0	0	1482
1479	14	67	0,0000	1414	1472	1485
1480	37	66	0,0000	0	0	1493
1481	57	64	0,0000	0	0	1485
1482	5	63	0,0000	1477	1478	1484
1483	58	62	0,0000	0	0	1484
1484	5	58	0,0000	1482	1483	1490
1485	14	57	0,0000	1479	1481	1497
1486	45	56	0,0000	1396	1420	1685
1487	42	54	0,0000	0	0	1490
1488	52	53	0,0000	1372	0	1747
1489	29	50	0,0000	0	0	1497
1490	5	42	0,0000	1484	1487	1494
1491	36	41	0,0000	0	0	1494
1492	31	38	0,0000	1471	0	1712
1493	15	37	0,0000	1204	1480	1707
1494	5	36	0,0000	1490	1491	1496
1495	30	33	0,0000	0	0	1496
1496	5	30	0,0000	1494	1495	1502
1497	14	29	0,0000	1485	1489	1747
1498	23	26	0,0000	0	0	1499
1499	22	23	0,0000	1470	1498	1757
1500	4	21	0,0000	1464	0	1716
1501	19	20	0,0000	0	0	1502
1502	5	19	0,0000	1496	1501	1504
1503	17	18	0,0000	0	0	1504
1504	5	17	0,0000	1502	1503	1506
1505	11	12	0,0000	0	0	1506
1506	5	11	0,0000	1504	1505	1508
1507	7	10	0,0000	0	0	1508
1508	5	7	0,0000	1506	1507	1766
1509	59	1499	0,0276	0	0	1532
1510	168	1047	0,0551	0	0	1558
1511	1472	1732	0,0827	0	0	1584
1512	161	1708	0,1102	0	0	1560
1513	1503	1583	0,1378	0	0	1559
1514	1121	1366	0,1653	0	0	1530
1515	1321	1352	0,1929	0	0	1516
1516	867	1321	0,2204	0	1515	1533
1517	598	1081	0,2480	0	0	1531
1518	93	1279	0,2757	0	0	1534
1519	1528	1679	0,3034	0	0	1562
1520	1208	1594	0,3312	0	0	1566
1521	1102	1581	0,3589	0	0	1563
1522	804	1137	0,3866	0	0	1564
1523	215	453	0,4143	0	0	1535
1524	1715	1721	0,4421	0	0	1582
1525	934	1666	0,4698	0	0	1617
1526	353	1636	0,4975	0	0	1536
1527	1033	1562	0,5252	0	0	1567
1528	943	1458	0,5530	0	0	1582
1529	965	1272	0,5807	0	0	1561

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1530	715	1121	0,6085	0	1514	1596
1531	598	978	0,6362	1517	0	1585
1532	59	1623	0,6640	1509	0	1577
1533	270	867	0,6918	0	1516	1668
1534	93	452	0,7196	1518	0	1565
1535	215	1184	0,7475	1523	0	1597
1536	353	601	0,7753	1526	0	1605
1537	1146	1363	0,8032	0	0	1538
1538	818	1146	0,8311	0	1537	1583
1539	1058	1746	0,8590	0	0	1556
1540	46	1647	0,8869	0	0	1541
1541	46	1589	0,9148	1540	0	1544
1542	1025	1243	0,9426	0	0	1568
1543	34	879	0,9705	0	0	1687
1544	46	589	0,9984	1541	0	1615
1545	778	1333	1,0352	111	0	1570
1546	156	274	1,0719	651	0	1571
1547	487	662	1,1089	0	753	1566
1548	196	610	1,1458	618	0	1586
1549	615	1388	1,1828	715	0	1587
1550	1268	1290	1,2198	0	372	1614
1551	470	1285	1,2567	304	0	1572
1552	95	1013	1,2937	821	0	1606
1553	294	796	1,3307	857	0	1598
1554	821	1397	1,3678	0	315	1619
1555	512	1050	1,4050	128	0	1727
1556	1058	1461	1,4465	1539	151	1569
1557	762	1273	1,4881	761	0	1602
1558	168	336	1,5298	1510	279	1588
1559	1231	1503	1,5714	180	1513	1663
1560	161	170	1,6131	1512	1010	1613
1561	233	965	1,6549	1103	1529	1633
1562	212	1528	1,6966	1162	1519	1601
1563	1102	1241	1,7384	1521	427	1631
1564	149	804	1,7801	843	1522	1637
1565	93	266	1,8243	1534	1027	1598
1566	487	1208	1,8706	1547	1520	1678
1567	397	1033	1,9203	538	1527	1622
1568	279	1025	1,9701	445	1542	1602
1569	739	1058	2,0207	814	1556	1628
1570	778	1411	2,0725	1545	83	1652
1571	156	775	2,1244	1546	842	1658
1572	470	515	2,1763	1551	329	1622
1573	137	165	2,2314	623	752	1609
1574	103	928	2,2868	1238	168	1599
1575	335	477	2,3423	536	1007	1609
1576	251	1266	2,3977	289	1	1618
1577	59	72	2,4534	1532	908	1616
1578	70	659	2,5091	471	132	1610
1579	269	1493	2,5649	1001	147	1620
1580	817	1376	2,6207	122	293	1612
1581	115	911	2,6765	58	337	1600
1582	943	1715	2,7323	1528	1524	1649
1583	461	818	2,7880	863	1538	1689
1584	228	1472	2,8505	1196	1511	1668
1585	71	598	2,9136	1304	1531	1682
1586	196	324	2,9779	1548	1121	1680
1587	615	764	3,0428	1549	882	1638
1588	139	168	3,1085	1057	1558	1701
1589	1053	1180	3,1746	413	542	1603
1590	136	392	3,2407	333	604	1604
1591	339	451	3,3073	858	1071	1634
1592	265	272	3,3738	309	1078	1621
1593	224	497	3,4403	938	596	1605
1594	209	1511	3,5073	1004	181	1635
1595	284	321	3,5742	357	1265	1653
1596	527	715	3,6438	611	1530	1682
1597	2	215	3,7134	1443	1535	1703
1598	93	294	3,7849	1565	1553	1700
1599	103	197	3,8565	1574	743	1653
1600	114	115	3,9282	969	1581	1662
1601	212	214	4,0049	1562	1200	1677
1602	279	762	4,0832	1568	1557	1656
1603	476	1053	4,1624	783	1589	1672
1604	136	700	4,2415	1590	437	1680
1605	224	353	4,3210	1593	1536	1662
1606	95	143	4,4023	1552	1243	1655
1607	49	697	4,4854	1227	331	1638
1608	178	219	4,5686	1351	1302	1636
1609	137	335	4,6521	1573	1575	1730
1610	24	70	4,7358	1229	1578	1702

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1611	84	1314	4,8194	1201	390	1629
1612	495	817	4,9031	816	1580	1630
1613	161	273	4,9882	1560	1063	1678
1614	51	1268	5,0747	1385	1550	1631
1615	35	46	5,1626	1446	1544	1714
1616	59	102	5,2518	1577	723	1721
1617	258	934	5,3442	1256	1525	1709
1618	251	597	5,4370	1576	1023	1657
1619	249	821	5,5299	1137	1554	1649
1620	65	269	5,6229	1431	1579	1651
1621	86	265	5,7162	1072	1592	1704
1622	397	470	5,8102	1567	1572	1687
1623	682	898	5,9052	587	746	1648
1624	252	328	6,0008	569	1267	1641
1625	172	309	6,0965	1383	1006	1639
1626	164	605	6,1921	509	850	1647
1627	315	623	6,2877	654	307	1640
1628	689	739	6,3847	974	1569	1719
1629	84	449	6,4851	1611	494	1686
1630	495	738	6,5873	1612	532	1713
1631	51	1102	6,6897	1614	1563	1663
1632	307	317	6,7937	1092	1052	1654
1633	173	233	6,8982	1328	1561	1669
1634	28	339	7,0039	1136	1591	1690
1635	55	209	7,1099	799	1594	1698
1636	178	415	7,2228	1608	1189	1645
1637	149	770	7,3379	1564	888	1705
1638	49	615	7,4541	1607	1587	1686
1639	8	172	7,5769	1405	1625	1707
1640	192	315	7,6997	1018	1627	1675
1641	252	381	7,8226	1624	942	1676
1642	244	535	7,9458	1273	301	1664
1643	43	290	8,0698	1455	1283	1660
1644	39	399	8,1937	629	848	1661
1645	178	260	8,3233	1636	1128	1722
1646	180	238	8,4572	822	1367	1684
1647	164	352	8,5922	1626	1251	1669
1648	401	682	8,7281	1120	1623	1671
1649	249	943	8,8652	1619	1582	1727
1650	68	144	9,0047	1474	849	1670
1651	65	358	9,1460	1620	1074	1677
1652	383	778	9,2888	1245	1570	1691
1653	103	284	9,4334	1599	1595	1701
1654	221	307	9,5785	1386	1632	1672
1655	3	95	9,7260	1326	1606	1689
1656	87	279	9,8740	1175	1602	1739
1657	145	251	10,0285	1246	1618	1710
1658	131	156	10,1920	1223	1571	1710
1659	422	1127	10,3556	1172	205	1713
1660	43	89	10,5210	1643	1415	1681
1661	39	275	10,6864	1644	1278	1702
1662	114	224	10,8545	1600	1605	1729
1663	51	1231	11,0228	1631	1559	1709
1664	175	244	11,2005	1022	1642	1711
1665	246	268	11,3796	1091	1346	1725
1666	25	232	11,5587	1454	1337	1688
1667	9	226	11,7389	1255	1114	1673
1668	228	270	11,9207	1584	1533	1695
1669	164	173	12,1050	1647	1633	1738
1670	68	363	12,2910	1650	1181	1703
1671	166	401	12,4789	1368	1648	1719
1672	221	476	12,6692	1654	1603	1742
1673	9	16	12,8599	1667	1467	1692
1674	91	148	13,0511	1428	690	1685
1675	1	192	13,2444	1444	1640	1683
1676	124	252	13,4376	1335	1641	1737
1677	65	212	13,6316	1651	1601	1750
1678	161	487	13,8310	1613	1566	1704
1679	48	746	14,0392	1418	900	1694
1680	136	196	14,2518	1604	1586	1740
1681	43	540	14,4670	1660	956	1736
1682	71	527	14,6933	1585	1596	1734
1683	1	76	14,9212	1675	1344	1746
1684	40	180	15,1506	1370	1646	1744
1685	45	91	15,3814	1486	1674	1721
1686	49	84	15,6121	1638	1629	1741
1687	34	397	15,8456	1543	1622	1728
1688	25	81	16,0800	1666	1395	1739
1689	3	461	16,3149	1655	1583	1724
1690	28	132	16,5587	1634	1373	1729
1691	283	383	16,8049	1333	1652	1734

