

UNIVERSIDADE FEDERAL DO PARANÁ

FLAVIANE PECCIN BREVI

MODELAGEM ESTATÍSTICA PARA GESTÃO DE CARTEIRA DE CLIENTES —
APLICAÇÃO EM TELECOMUNICAÇÕES



CURITIBA
2012

FLAVIANE PECCIN BREVI

**MODELAGEM ESTATÍSTICA PARA GESTÃO DE CARTEIRA DE CLIENTES
— APLICAÇÃO EM TELECOMUNICAÇÕES**

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Métodos Numéricos em Engenharia na Área de Concentração em Programação Matemática e na Linha de Pesquisa em Métodos Estatísticos Aplicados à Engenharia, como requisito parcial à obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Anselmo Chaves Neto

CURITIBA
2012

TERMO DE APROVAÇÃO

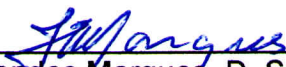
FLAVIANE PECCIN

MODELAGEM ESTATÍSTICA PARA GESTÃO DE CARTEIRA DE CLIENTES — APLICAÇÃO EM TELECOMUNICAÇÕES

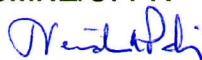
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Ciências, no Curso de Pós-graduação em Métodos Numéricos em Engenharia na Área de Concentração em Programação Matemática e na Linha de Pesquisa em Métodos Estatísticos Aplicados à Engenharia, dos Setores de Tecnologia e de Ciências Exatas da Universidade Federal do Paraná, pela seguinte banca examinadora:



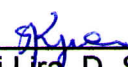
Anselmo Chaves Neto, D. Sc.
Orientador — PPGMNE/UFPR



Jair Mendes Marques, D. Sc.
PPGMNE/UFPR



Neida Maria Patias Volpi, D. Eng.
PPGMNE/UFPR



Sachiko Araki Lira, D. Sc.
PGMEC/UFPR

Curitiba, 05 de outubro 2012.

Dedico este trabalho ao meu esposo, pela sua compreensão em meus momentos de ausência e pelo apoio nas horas difíceis, permitindo-me chegar ao fim de mais esta importante etapa em minha vida.

AGRADECIMENTOS

Agradeço a Deus por me proporcionar esta oportunidade e me permitir chegar até aqui, concedendo-me forças e dedicação para conciliar os estudos do mestrado com a vida profissional e ainda encontrar algum tempo para família e amigos.

Agradeço em especial ao Professor Dr. Anselmo Chaves Neto, por confiar no meu potencial, pelos incontáveis momentos em que me incentivou e por compartilhar seus conhecimentos.

Ao meu esposo Angelo, por sua paciência, carinho e suporte durante essa fase e por me fazer entender o significado de algumas palavras, como companheirismo e cumplicidade.

À Universidade Federal do Paraná, especialmente ao Departamento de Estatística, e ao Programa de Pós-graduação em Métodos Numéricos em Engenharia, pela oportunidade que me proporcionaram de aperfeiçoar meus estudos.

Aos professores das disciplinas cursadas.

Aos colegas, que contribuíram sempre que possível.

RESUMO

O setor de telecomunicações brasileiro entrou num processo contínuo de expansão após as privatizações que ocorreram na década de 1990. Com esse processo, entraram nesse mercado novos competidores e novas tecnologias. Isso foi benéfico para o consumidor, que atualmente tem mais opções de escolha de prestadoras do serviço. Por outro lado, as empresas que oferecem esse serviço precisam quantificar as probabilidades de esses clientes pagarem pelo serviço prestado. Nesse contexto, este trabalho tem por objetivos mapear as variáveis que podem ser consideradas fatores de risco para inadimplência, construir um modelo preditivo para estimar a probabilidade de pagamento das faturas mensais e avaliar a possibilidade da implantação desse modelo num sistema automático de decisão. Para o desenvolvimento dos modelos, foram considerados 43.339 clientes da empresa em estudo, que contrataram os serviços entre janeiro e dezembro de 2007. Esses clientes foram acompanhados por doze meses para avaliar sua performance quanto ao pagamento de suas faturas mensais. Dessa forma, a amostra foi dividida em dois grupos, denominados *bons* e *maus*. Clientes *bons* são aqueles que, durante os doze meses, atingiram no máximo cinco dias em atraso, no total de 23.095 clientes. *Maus* são aqueles que atingiram 90 dias ou mais de atraso para pagamento de pelo menos uma fatura no período observado, no total de 20.244 clientes. Os clientes que tiveram atraso maior que cinco dias em alguma das faturas, mas não atingiram noventa dias de atraso foram classificados como indeterminados e não fazem parte da amostra em estudo. Com o objetivo de segmentar a amostra em grupos homogêneos, antes de desenvolver o modelo preditivo, foi aplicada a Análise de *Cluster*. Por meio dessa metodologia, não foi possível obter agrupamentos de clientes. Isso se deve à homogeneidade do grupo em estudo, que gera alta concentração num único *cluster*. Para obter os modelos preditivos que estimam a probabilidade de pagamento, foi utilizada a Regressão Logística. As variáveis preditoras foram consideradas no modelo por duas formas: a primeira utilizando variáveis *dummies*; a segunda, o WOE (*Weight of Evidence*). Em ambas as formas de ajuste do modelo, foram identificados fatores de risco semelhantes, tais como parcelamento, alertas de fraude, altos valores de fatura e baixos históricos de pagamentos, clientes novos, tipo de produto contratado e perfil de consumo. Além desses, algumas regiões podem ser apontadas como fatores de risco, identificadas por meio de CEP, cidade e estado.

Palavras-chave: Telecomunicações; Inadimplência; Análise de *Cluster*; Regressão Logística.

ABSTRACT

The Brazilian sector of telecommunications witnessed a streak of continuous expansion after becoming privatized in the 1990's. During this process many new competitors and new forms of technology were introduced to the market. This expansion benefitted the consumer by offering more options and services to choose from. On the other hand, the companies that offer these services need to analyze the possibility that the clients will pay for the service being given. Given this fact, the companies must chart the variables that can be considered factors which indicate risk of default, and construct a predictive model to estimate the probability of monthly payments being made along with assessing the possibility of implementing this model in an automatic decision-making process. 43.339 clients of the company which contracted the services of the company between January and December of 2007 were considered when developing this model. These clients were tracked for twelve months to assess their tendencies related to their monthly payments. This sample was divided into two groups named Good and Bad. Good clients were those who during the twelve months had a maximum of five late days in making their payments. The number of clients in this group was 23.095. Bad clients were those who had total of 90 late days in making their payments. The number of clients in this group was 20.244. A Cluster Analysis was used with the purpose of segmenting the sample before developing the predictive model. By using this methodology it was not possible to group these clients. This was due to the group homogeneity in the study that created a high concentration while using only one cluster. Logistic Regression was used to obtain predictive models that estimate the possibility of payment. The predictable variables were considered in the model by using two forms. The first form used dummies variables and the second form used WOE (Weight of Evidence) variables. In both forms for the model, similar risk factors were identified, such as division of payments, fraud alert, high and low payment values, new clients, the type of product purchased, and the customer profile. Besides these factors, some geographic regions can be seen as contributing to risk. These regions can be identified by tracking area codes, cities and states.

Keywords: Telecommunications, Default, Cluster Analysis, Logistic Regression

LISTA DE FIGURAS

Figura 1 — Dendrograma — Método do centroide.....	50
Figura 2 — Dendrograma — Método das médias das distâncias	51
Figura 3 — Dendrograma — Método de Ward.....	52
Figura 4 — Distribuição acumulada para os clientes Bons e Maus pela probabilidade de pagamento da fatura — Modelo com variáveis <i>dummies</i>	64
Figura 5 — Distribuição acumulada para os clientes Bons e Maus pela probabilidade de pagamento da fatura — Modelo WOE	68

LISTA DE QUADROS

Quadro 1 — Matriz de confusão genérica.....	23
Quadro 2 — Valores referenciais para KS	25
Quadro 3 — Tabela de contingência 2 X 2	36
Quadro 4 — Matriz de Classificação — Lachenbruch.....	45
Quadro 5 — Variáveis analisadas	47
Quadro 6 — Critério para classificação das categorias das variáveis.....	54
Quadro 7 — Modelo ajustado com variáveis <i>dummies</i> — com intercepto	58
Quadro 8 — Análise do Desvio - Modelo com variáveis <i>dummies</i> — com intercepto	60
Quadro 9 — Modelo ajustado com variáveis <i>dummies</i> — sem intercepto	60
Quadro 10 — Análise do Desvio - Modelo com variáveis <i>dummies</i> — sem intercepto ..	63
Quadro 11 — Modelo ajustado usando WOE	64
Quadro 12 — Análise do Desvio - Modelo WOE.....	67

LISTA DE TABELAS

Tabela 1 — Evolução anual do número de acessos no Brasil (milhões).....	15
Tabela 2 — Usuários de internet no Brasil, de acordo com várias fontes	15
Tabela 3 — Quantidade de domicílios com internet (milhões)	16
Tabela 4 — ANODEV — Deviance Analysis	42
Tabela 5 — Distribuição dos clientes por mês de contratação.....	46
Tabela 6 — Intervalo de confiança para a proporção de clientes bons.....	53
Tabela 7 — Fusão das categorias da variável estado.....	54
Tabela 8 — Análise bivariada — Estado conforme os grupos de risco.....	55
Tabela 9 — Reagrupamento das categorias da variável estado	55
Tabela 10 — Discretização da variável valor da fatura	56
Tabela 11 — Discretização final — variável valor da fatura	57
Tabela 12 — Matriz de confusão para o modelo com variáveis <i>dummies</i> — com intercepto — Amostra de desenvolvimento	68
Tabela 13 — Matriz de confusão para o modelo com variáveis <i>dummies</i> — sem intercepto — Amostra de desenvolvimento	69
Tabela 14 — Matriz de confusão para o modelo usando WOE — Amostra de desenvolvimento	69
Tabela 15 — Comparação dos resultados dos modelos na amostra de desenvolvimento	69
Tabela 16 — Matriz de confusão para o modelo usando variáveis <i>dummies</i> — Com intercepto — Obtido por Lachenbruch.....	70
Tabela 17 — Matriz de confusão para o modelo usando variáveis <i>dummies</i> — Sem intercepto — Obtido por Lachenbruch.....	70
Tabela 18 — Matriz de confusão para o modelo usando WOE — Obtido por Lachenbruch.....	71

SUMÁRIO

1. INTRODUÇÃO	12
1.1. TEMA DO ESTUDO	12
1.2. OBJETIVOS	12
1.3. JUSTIFICATIVA	13
1.4. ESTRUTURA DO TRABALHO	13
2. REVISÃO DE LITERATURA	14
2.1. O SETOR DE TELECOMUNICAÇÕES	14
2.2. GESTÃO DE RISCOS DE CRÉDITO MASSIFICADO	16
2.2.1. Modelos de <i>application</i> e <i>behaviour score</i>	17
2.2.2. Modelos julgamentais e quantitativos	17
2.3. ROTEIRO PARA O DESENVOLVIMENTO DE UM MODELO DE <i>SCORING</i> ..	18
2.4. TÉCNICAS ESTATÍSTICAS	26
2.4.1. Análise de <i>Cluster</i>	26
2.4.2. Regressão Logística	32
2.4.3. Variável <i>dummy</i>	35
2.4.4. Razão de chances: <i>odds ratio</i>	35
2.4.5. Estimadores de máxima verossimilhança	36
2.4.6. Seleção passo a passo: <i>stepwise</i>	38
2.4.7. Significância das covariáveis no modelo	41
2.4.8. Diagnóstico do modelo ajustado	43
2.4.9. Avaliação do erro do modelo ajustado — Método de Lachenbruch	44
3. MATERIAL E MÉTODOS	46
3.1. MATERIAL	46
3.1.1. Amostra	46
3.1.2. Definição de bom e mau	47
3.1.3. Variáveis	47
3.2. METODOLOGIA E PROCEDIMENTOS	49
4. RESULTADOS E DISCUSSÃO	50
4.1. ANÁLISE DE <i>CLUSTER</i>	50
4.2. TRATAMENTO DAS VARIÁVEIS CATEGÓRICAS	53
4.3. TRATAMENTO DAS VARIÁVEIS CONTÍNUAS (DISCRETIZAÇÃO)	55

4.4. AJUSTE DO MODELO USANDO VARIÁVEIS <i>DUMMIES</i>	57
4.5. AJUSTE DO MODELO USANDO O WOE COMO VARIÁVEL PREDITORA....	64
4.6. COMPARAÇÃO DOS MODELOS NA AMOSTRA DE DESENVOLVIMENTO .	68
4.7. VALIDAÇÃO DO MODELO PELO MÉTODO DE LACHENBRUCH.....	70
5. CONSIDERAÇÕES FINAIS E CONCLUSÃO	72
REFERÊNCIAS	74
APÊNDICES	76

1. INTRODUÇÃO

1.1. TEMA DO ESTUDO

Após as privatizações ocorridas na década de 90, o setor de telecomunicações brasileiro entrou em contínua expansão. A entrada de novos competidores e novas tecnologias levou a um crescimento das expectativas do cliente. Isso está causando maiores taxas de *churn*, ou melhor, de taxas de atrito com os clientes, bem como maiores custos de aquisição de clientes novos. Para minimizar esse problema, as ações a serem tomadas devem levar em conta o valor de cada cliente. Dentro desse contexto, a segmentação da base de clientes é o primeiro passo para as empresas operadoras desse serviço.

O tema objeto deste projeto é importante devido ao aumento da concorrência e do nível de exigência dos clientes, de modo que as empresas que prestam esse serviço precisam tratar seus clientes de forma diferenciada, visando a fidelizar e manter em sua carteira aqueles mais rentáveis. Além disso, estimar a probabilidade de pagamento é importante para os planos estratégicos e financeiros da empresa, pois o pagamento das faturas influencia diretamente seu fluxo de caixa.

1.2. OBJETIVOS

O objetivo geral desse trabalho é aplicar um modelo estatístico a uma carteira de clientes de uma empresa do setor de telecomunicações.

Os objetivos específicos deste trabalho são:

- a) Mapear as variáveis preditoras de inadimplência;
- b) Desenvolver modelos de *behaviour scoring* utilizando técnicas estatísticas e computacionais;
- c) Comparar os modelos desenvolvidos quanto à qualidade de predição e aplicabilidade ao negócio;
- d) Avaliar a possibilidade de implantação do modelo escolhido num sistema automático de decisão.

1.3. JUSTIFICATIVA

Sempre que um cliente adquire um produto ou serviço sem pagar imediatamente por ele, há um risco associado para o fornecedor do produto ou serviço. No processo de concessão de serviço de telecomunicações, a empresa aceita esse risco de inadimplência para cada novo cliente. Além disso, todos os clientes novos ou antigos geram mensalmente uma fatura a ser paga em data pré-estabelecida. Mensurar o risco de pagamento (ou não pagamento) é de vital importância nesse setor, pois um cliente mal classificado pode causar prejuízos.

1.4. ESTRUTURA DO TRABALHO

Este trabalho é composto por essa introdução, por uma revisão de literatura que está no capítulo 2 e na qual se aborda desde o setor de telecomunicações até as técnicas estatísticas que foram aplicadas no trabalho. A seguir, na parte de material e métodos é descrito o banco de dados utilizado, bem como as metodologias aplicadas. No capítulo 4, são apresentados os resultados e é feita uma discussão sobre eles. Finalmente, apresenta-se a conclusão, no capítulo 5, com as recomendações finais. No final têm-se, ainda as referências citadas no trabalho e quatro apêndices. No Apêndice um apresenta-se o tratamento das variáveis categóricas. No Apêndice dois são apresentadas as análises bivariadas. No Apêndice três tem-se a tabela final com todos os agrupamentos das categorias das variáveis e, por último, no Apêndice quatro é apresentado o tratamento das variáveis contínuas.

2. REVISÃO DE LITERATURA

2.1. O SETOR DE TELECOMUNICAÇÕES

Segundo o Idec (Instituto Brasileiro de Defesa do Consumidor), o serviço de telefonia fixa é considerado como essencial e deve ser prestado a todos os brasileiros. Com a privatização, o controle da rede de telefonia e os investimentos passaram a ser de responsabilidade das empresas privadas, mas a infraestrutura é reversível à união. Para concluir o processo de privatização das telecomunicações e organizar a exploração e a prestação do serviço de telefonia fixa, o governo federal dividiu o Brasil em quatro regiões em 1998. Em cada uma das quatro regiões criadas pelo PGO (Plano Geral de Outorgas), uma empresa — chamada de **concessionária** — tem a obrigação de garantir a universalização e a continuidade do serviço de telefonia fixa. No entanto, para que essas empresas não sejam as únicas atuando em suas regiões de concessão, ou seja, não constituam monopólio, deixando o consumidor refém de um único prestador, passaram a existir as empresas chamadas de **autorizadas**, que prestam serviço de telefonia fixa em regime privado e não têm obrigação de garantir a universalização e a continuidade do serviço.

Os acessos telefônicos fixos são classificados em:

- a) **acessos instalados** — conjunto formado por todos os acessos fixos que disponham de todas as facilidades necessárias para estar em serviço.
- b) **acessos em serviço** — acessos instalados que estão colocados à disposição dos usuários, inclusive os terminais de uso público (TUP).

A Tabela 1 apresenta a evolução do número de acessos fixos instalados e em serviço no Brasil. Ela inclui acessos das concessionárias e das novas autorizações de STFC (Serviço Telefônico Fixo Comutado).

Tabela 1 — Evolução anual do número de acessos no Brasil (milhões)

Ano	Acessos instalados	Acessos em serviço	Acessos em serviço/100 habitantes
2010	62,0	42,0	21,7
2009	59,6	41,5	21,6
2008	57,9	41,2	21,7
2007	52,7	39,4	20,9
2006	51,2	38,8	20,8
2005	50,5	39,8	21,6
2004	50,0	39,6	21,7
2003	49,8	39,2	21,8
2002	49,2	38,8	21,9
2001	47,8	37,4	21,4
2000	38,3	30,9	17,9
1999	27,8	25,0	14,7
1998	22,1	20,0	12,0
1997	18,8	17,0	10,3
1996	16,5	14,8	9,1
1995	14,6	13,3	8,3
1994	13,3	12,3	7,8

Fonte: Site Teleco (Teleco - Inteligência em Telecomunicações)

Todos os clientes analisados neste trabalho possuem acesso à telefonia fixa, e parte deles tem acesso à internet. Apresenta-se a seguir a quantidade de usuários de internet no Brasil, de acordo com várias fontes.

Tabela 2 — Usuários de internet no Brasil, de acordo com várias fontes

Milhões	2005	2006	2007	2008	2009
Fonte: PNAD ⁽¹⁾	32,1	55,9	67,9
Fonte: TIC Domicílios ⁽¹⁾	...	35,3	44,9	53,9	63,0
Fonte Ibope ⁽²⁾	...	32,5	39	62,3	66,3

Fonte: Site Teleco (Teleco - Inteligência em Telecomunicações)

(1) população de 10 anos de idade ou mais que acessou a internet pelo menos uma vez nos 90 dias que antecederam a entrevista.

(2) total de pessoas com mais de 16 anos com acesso à internet em qualquer ambiente.

Tabela 3 — Quantidade de domicílios com internet (milhões)

Período	2006	2007	2008	2009	2010	2011
Jan	867	1.037	1.241	1.553	1.965	2.350
Fev	881	1.053	1.251	1.572	1.986	2.368
Mar	899	1.074	1.274	1.608	2.023	2.410
Abr	912	1.090	1.298	1.653	2.065	2.462
Mai	928	1.114	1.342	1.691	2.107	2.499
Jun	944	1.135	1.375	1.725	2.135	2.545
Jul	958	1.152	1.416	1.771	2.172	2.590
Ago	978	1.174	1.449	1.808	2.204	2.625
Set	991	1.193	1.473	1.844	2.236	2.655
Out	1.006	1.212	1.498	1.904	2.275	...
Nov	1.021	1.226	1.527	1.941	2.299	...
Dez	1.025	1.231	1.534	1.953	2.318	...

Fonte: Site Teleco (Teleco - Inteligência em Telecomunicações)

2.2. GESTÃO DE RISCOS DE CRÉDITO MASSIFICADO

Conforme Manfio (2007), ao se analisar o risco de crédito, tem-se que abordar grande quantidade de eventos, por exemplo, empréstimos, seguros, transações com cartão de crédito ou débito. Outro ponto de destaque é que esses eventos, em sua maioria, são realizados por pessoas físicas. Há relação entre o risco de perda e a inadimplência. Sendo assim, mesmo com juros percentualmente altos, quando aplicados a valores relativamente baixos, resultam em pequenos valores, ou seja, pequenas receitas por unidade e somente fazem a diferença se a quantidade de unidades for grande. Da mesma forma, o risco percentualmente alto aplicado a baixos valores de empréstimos resulta em valores de perda baixos por evento. Portanto, o que faz a diferença é a grande quantidade. Com isso, de acordo com Manfio (2007) conclui-se que os processos operacionais devem ser de custo baixo e, sempre que possível, automatizados. Em resumo, as características desse universo são: grandes quantidades, pequenos valores de empréstimo, juros e riscos percentualmente altos, receita unitária pequena, custo baixo, agilidade alta e alto nível de automação.

2.2.1. Modelos de *application* e *behaviour score*

Conforme mencionado em Sicsú (2010), os modelos de *credit scoring* podem ser aplicados tanto a clientes com os quais o credor não teve experiência anterior quanto a solicitantes que já foram seus clientes. Os modelos desenvolvidos para novos clientes são chamados de *application scoring*. Os modelos utilizados para clientes ou ex-clientes de crédito do credor são chamados de *behaviour score* e utilizam, além das informações consideradas no *application scoring*, as informações relativas a créditos anteriores. Por exemplo, se o cliente pagou em dia eventuais parcelas anteriores, essa informação contribuirá de forma positiva no cálculo do *score*. Em contrapartida, se o cliente atrasou várias parcelas anteriores, seu *score* será reduzido em razão dessa experiência negativa. Em função da incorporação dessas informações, os modelos de *behaviour scoring*, alvo deste estudo, tendem a fornecer maior poder de discriminação que os modelos de *application scoring*.

2.2.2. Modelos julgamentais e quantitativos

Segundo Thomas (2000), a metodologia utilizada na construção de modelos de *credit scoring* era, originalmente, julgamental. Nos modelos julgamentais, as variáveis que compõem o *score* e seus respectivos pesos são determinadas pelos gestores de crédito da instituição com base em critérios subjetivos. Conforme Sicsú (2010), esses modelos somente são indicados como última opção quando não se dispõe de dados suficientes para desenvolver um modelo estatístico, pois não permitem quantificar a probabilidade de perda, que é o elemento mais importante na avaliação do risco de crédito. Andrade (2004) ressalta que, embora algumas instituições ainda utilizem modelos de *credit scoring* julgamentais, atualmente a vasta maioria utiliza os modelos quantitativos, construídos a partir de técnicas de análise estatística multivariada, tais como a análise discriminante e a regressão logística, ou modelos de inteligência artificial, como redes neurais.

Em Sicsú (2010), os modelos quantitativos são divididos em generalistas e customizados. Os generalistas são aqueles disponibilizados por empresas de informações comerciais e estimam o risco de crédito de um indivíduo sem considerar operação específica e o mercado específico. Esses modelos são uma solução interessante para produtos novos ou novos mercados sobre os quais o credor não possui base de dados. Os modelos customizados são os mais adequados para avaliação do risco de crédito, pois são desenvolvidos levando-se em consideração o mercado específico do credor, as características da operação para a qual o modelo será usado e a definição de perda específica da operação. Sendo assim, esses modelos permitem maior acurácia na estimação do risco de crédito.

2.3. ROTEIRO PARA O DESENVOLVIMENTO DE UM MODELO DE *SCORING*

Segundo Sicsú (2010), as principais etapas para o desenvolvimento de um modelo de *scoring* são:

1. Planejamento e definições: nessa fase deve-se definir o escopo do projeto, os tipos de objetivos que se esperam alcançar, os tipos de operações de crédito a que se destinam o modelo e qual o mercado de atuação. Nessa fase também é necessário levantar informações sobre as normas que regulamentam a concessão e a gestão da operação a que o modelo será aplicado. Outro item importante nessa fase é a definição de possíveis segmentações do mercado-alvo. É comum que essas segmentações sejam definidas de forma subjetiva, como por regiões geográficas. A segmentação pode também ser obtida por meio da aplicação de técnicas estatísticas específicas para agrupamento de indivíduos similares, denominada Análise de *Clusters*. Uma definição muito importante a ser analisada nessa fase é “definição de bom/mau”. Essa é provavelmente a definição operacional mais complexa, pois o cliente denominado “mau” é o cliente que causa perdas inaceitáveis à instituição credora. Clientes não classificados dessa forma serão classificados como “bons”, “intermediários” ou “indefinidos” ou podem não ser classificados. Portanto, cada instituição precisa definir de forma clara o que considera perda, pois essa definição

representará claramente o objetivo para o qual o modelo foi desenvolvido. As definições de bom/mau podem ser restritas ao comportamento do cliente exclusivamente em seu relacionamento com o credor (performance interna) ou considerar também informações de mercado (performance externa). O desenvolvimento do modelo é baseado na experiência com os créditos concedidos no passado. Essa experiência é estudada obtendo-se amostra de créditos concedidos no passado e classificando-os de acordo com as definições de bom/mau. Os créditos que compõem a amostra devem ser suficientemente atualizados para refletir o perfil dos clientes atuais e suficientemente antigos, para que se possa observar sua performance quanto ao pagamento. Para definir o período de concessão (período durante o qual os créditos foram concedidos), é necessário, além de garantir que se tenha amostra representativa, que ela contemple eventuais sazonalidades do processo. Para que se possa classificar os tomadores desses créditos como bom/mau, é preciso observá-los durante um período de pagamentos, denominado período de performance. O histórico do cliente anterior à data de concessão de crédito atual é importante para prever se ele se tornará um mau pagador. Verifica-se no histórico se o solicitante apresentou problemas de crédito na instituição credora ou em outros agentes financeiros (por meio de informações de *bureau* de crédito). Esse período é denominado período histórico.

2. Identificação das variáveis potenciais: a eficácia do modelo depende diretamente das informações utilizadas. Inicialmente, identifica-se um conjunto de variáveis que possa ser utilizado para discriminar clientes bons dos clientes maus. Essas variáveis podem ser informações sociodemográficas, financeiras, de relacionamento com o credor e comportamentais (relativas a créditos históricos). Após a identificação das variáveis preditoras, deve-se verificar a confiabilidade de cada uma, a disponibilidade ao longo do tempo, os aspectos éticos e legais para sua utilização.

3. Planejamento e seleção da amostra: o modelo é desenvolvido com base em amostra de clientes que tomaram crédito no passado. Portanto, é necessário um plano amostral para selecionar esses clientes. Hoje, devido aos recursos computacionais existentes, muitos analistas trabalham com toda a base de dados disponível na instituição credora. Para selecionar uma amostra para desenvolver o modelo, costuma-se recorrer a duas técnicas: amostragem aleatória simples e

amostragem aleatória estratificada. Para dimensionar a amostra, alguns autores indicam que ela deve ser igual ou dez vezes maior que o número de variáveis. Essa regra não tem justificativa conhecida e pode conduzir a amostras insatisfatórias. Sempre que possível, a amostra deve ser suficientemente grande, para que se possa dividi-la em duas partes; uma para desenvolvimento do modelo e outra para validação. A separação entre desenvolvimento e validação é feita de forma aleatória. Em geral, de 50% a 70% são utilizados para desenvolvimento do modelo, e o restante para validação. Essa divisão é feita porque, ao se avaliar o modelo na mesma amostra de desenvolvimento, os resultados podem ser superestimados, portanto é importante validá-lo numa amostra específica para esse fim.

4. Análise e tratamento dos dados: é importante, antes de aplicar as técnicas estatísticas, verificar se os dados foram coletados de forma correta, analisar as características de cada variável individualmente e entender a relação entre essas variáveis.

a) Análise univariada: é útil para entender o comportamento de cada variável. Pode-se, por exemplo, fazer distribuição de frequência. Nessa fase, os valores inconsistentes são identificados e tratados e é verificada a existência de *missings*, *outliers* e o excesso de valores cadastrados como “outros”.

b) Discretização de variáveis: consiste em dividir a variável quantitativa em faixas de valores (categorias). Alguns autores utilizam o termo categorizar quando se referem à discretização de variáveis quantitativas. Há várias razões que motivam a discretização de variáveis:

- i. Permite entender com mais facilidade a relação dessa variável com o status do cliente (bom/mau), ou seja, quais categorias da variável representam maior risco de crédito, quais são neutras e quais têm o menor risco.
- ii. Para implementação do modelo, é necessário convencer as áreas de negócio, e os analistas dessas áreas compreendem mais facilmente o comportamento das variáveis quando apresentadas em categorias.
- iii. No caso de variável ordenada (qualitativa ou quantitativa), a discretização permite analisar se a variação do risco também segue essa ordenação.

- iv. Quando a relação entre a variável quantitativa e a medida do risco não é monotônica, a discretização é vantajosa, pois cada categoria é tratada de forma independente, criando-se variáveis *dummies* para elas.
- v. A discretização de variáveis quantitativas pode mitigar o efeito de fortes assimetrias gerando-se uma categoria que contenha todos os valores acima de determinado número.

Há também desvantagens ao se fazer a discretização, pois se perde informação ao agrupar indivíduos com valores distintos e tratá-los de forma semelhante. Existem diversos métodos para se fazer a discretização de variáveis. A primeira recomendação é que o número de indivíduos em cada categoria não seja pequeno. Uma regra comum entre os analistas é que cada classe não tenha frequência inferior a 5% da amostra total. Com as classes definidas, e considerando-se o tamanho mínimo, algumas classes podem ser agrupadas. Essa fusão de categorias pode ser feita com base na experiência do analista ou em algum critério estatístico. É possível também adotar um misto desses dois procedimentos. Usualmente, a fusão das categorias é baseada nas porcentagens de clientes bons e maus em cada categoria.

c) Geração das variáveis *dummies*: definidas as categorias, criam-se as variáveis *dummies*. Uma variável *dummy* é aquela que assume apenas dois valores — 0 (condição não é atendida) ou 1 (condição é atendida). Se uma variável apresenta k categorias, gera-se $k-1$ variáveis *dummies*.

d) Análises bivariadas: úteis para entender a relação de cada variável preditora com a variável *status* (bom/mau). Têm como principais objetivos analisar o potencial de discriminação de cada variável preditora e refinar a discretização dessas variáveis.

Diferentes critérios podem ser utilizados para analisar o poder de discriminação de cada categoria e decidir quais serão agrupadas, tais como:

- i. Risco relativo (RR): fornece o risco de desenvolvimento de determinada condição para um grupo comparado a outro grupo (Giolo,2006). Essa medida é expressa por:

$$RR = \frac{p_{11}}{p_{12}} \quad (1)$$

Em que p_{11} é a probabilidade de qualquer indivíduo pertencente à primeira coluna ser classificado na primeira linha e, analogamente, p_{12} é a probabilidade de qualquer indivíduo pertencente à segunda coluna ser classificado na primeira linha. Nesse caso, tem-se $\sum_{i=1}^k p_{i1} = 1$ e $\sum_{i=1}^k p_{i2} = 1$.

Um intervalo de confiança para RR é obtido por meio da estatística f^* , obtida tomando-se o logaritmo neperiano do RR, ou seja:

$$f^* = \ln(RR) = \ln(p_{11}) - \ln(p_{12}) \quad (2)$$

E a variância dessa estatística é:

$$V(f^*) = \frac{(1-p_{11})}{(n_{1+})(p_{11})} + \frac{(1-p_{12})}{(n_{2+})(p_{12})} \quad (3)$$

Para a qual n_{1+} é o total de indivíduos da coluna 1 e n_{2+} é o número total de indivíduos da coluna 2.

Portanto, o intervalo de confiança para o logaritmo neperiano do RR, ao nível de $100*(1-\alpha)\%$ de confiança, é dado por:

$$\left(f^* - Z_{\alpha/2} * \sqrt{V(f^*)} \leq \ln(RR) \leq f^* + Z_{\alpha/2} * \sqrt{V(f^*)} \right) = 1-\alpha \quad (4)$$

em que $Z_{\alpha/2}$ denota o percentil da distribuição normal padrão correspondente ao nível de confiança fixado.

ii. **Peso da evidência (WOE — *Weight of Evidence*)**

O WOE é uma forma de medir a distância entre as categorias das duas distribuições (distribuição de bons e distribuição de maus).

A expressão do WOE numa categoria da variável preditora é:

$$WOE = \ln\left(\frac{p_{11}}{p_{12}}\right) = \ln(RR) \quad (5)$$

iii. **A estatística IV (*Information Value*):**

Essa estatística permite avaliar o poder de discriminação de uma variável. Se as distribuições de bom/mau forem realmente diferentes, a variável permitirá discriminar esses dois tipos de cliente.

Essa estatística IV corresponde à soma dos WOE calculados e ponderados após o agrupamento das categorias. É definida pela expressão:

$$IV = \sum_{i=1}^k [p_{i1} - p_{i2}] * WOE_i \quad (6)$$

Em que p_{i1} é a probabilidade de qualquer indivíduo pertencente à primeira coluna ser classificado na linha i , e, analogamente, p_{i2} é a probabilidade de qualquer indivíduo pertencente à segunda coluna ser classificado na linha i .

5. Cálculo da fórmula de escoragem:

Existem várias metodologias para obter a fórmula de escoragem. A mais utilizada é a regressão logística, que permite estimar a probabilidade de ocorrência de um evento a partir de um conjunto de informações que caracteriza esse evento.

6. Análise e validação da fórmula:

A validação do modelo é feita previamente, de forma subjetiva, verificando se as variáveis e seus pesos fazem sentido. Uma validação mais detalhada é feita seguindo critérios estatísticos para avaliar o poder de discriminação do modelo. Podem-se destacar os seguintes indicadores:

i. Matriz de confusão

Dado que se tem a classificação dos clientes como bons ou maus, e que o modelo para classificação já foi ajustado, pode-se comparar a classificação observada (bom/mau) com a classificação obtida pelo modelo, conforme ilustrado no quadro a seguir:

Quadro 1 — Matriz de confusão genérica

Amostra observada	Amostra classificada pelo modelo		Total
	1	2	
1	n_{11}	n_{12}	N_1
2	n_{21}	n_{22}	N_2

onde n_{ij} o número de elementos pertencentes à população de origem i e que são classificados pelo modelo como pertencentes à população j . Quando $i = j$, tem-se o número de classificações corretas, e quando $i \neq j$, tem-se o número de classificações incorretas.

ii. Taxa aparente de erro

Com base na matriz de confusão, as estimativas das probabilidades de ocorrência de erros de classificação são dados respectivamente por:

$$\hat{p}(2/1) = \frac{n_{12}}{n_1} \quad \text{e} \quad \hat{p}(1/2) = \frac{n_{21}}{n_2} \quad \text{e} \quad \hat{E}(APER) = \frac{n_{12} + n_{21}}{n_1 + n_2}$$

Onde:

$\hat{p}(2/1)$ é a probabilidade estimada de classificar erroneamente um elemento em 2 quando ele é de 1;

$\hat{p}(1/2)$ é a probabilidade estimada de classificar erroneamente um elemento em 1 quando ele é de 2.

$\hat{E}(APER)$ é a taxa aparente de erro estimada.

iii. KS — índice de Kolmogorov-Smirnov

Segundo Sicsú (2010), a estatística KS é uma medida popular entre os analistas de risco para avaliar o desempenho dos modelos. Essa medida é a maior distância entre as funções de distribuição acumulada das pontuações dos clientes bons e maus, $F_b(k)$ e $F_m(k)$, respectivamente.

$$F_b(k) = \frac{\text{número de bons com pontuação} \leq k}{\text{número de bons}}$$

$$F_m(k) = \frac{\text{número de maus com pontuação} \leq k}{\text{número de maus}}$$

Em que k varre o conjunto de possíveis valores de pontuação. Calcula-se os valores dessas funções para k variando do menor ao maior valor possível de pontuação e depois obtém-se a maior diferença entre essas funções.

$$KS = \text{máx} [F_m(k) - F_b(k)]$$

Em geral, essa medida é expressa em porcentagem e seus valores de referência são apresentados no Quadro 2.

Quadro 2 — Valores referenciais para KS

Valores de KS	<i>Credit Scoring</i>	<i>Behavioral Scoring</i>
KS < 20%	Baixo	Baixo
20% < KS ≤ 30%	Baixo/aceitável	Baixo
30% < KS ≤ 40%	Bom	Baixo
40% < KS ≤ 50%	Muito Bom	Aceitável
50% < KS ≤ 60%	Excelente	Bom
60% < KS ≤ 70%	Valores pouco usuais	Muito Bom
KS > 70%		Excelente, valores pouco usuais

Fonte : Sicsú (2010)

iv. Coeficiente de Gini

Conforme Sicsú (2010), o coeficiente de Gini é de certa forma equivalente a curva ROC. É o dobro da área entre a curva ROC e a linha diagonal correspondente a classificação aleatória.

7. Ajuste final do modelo:

O desenvolvimento das análises do modelo se baseia nos clientes que contrataram o crédito. Isso causa um viés quando o interesse é aplicar esse modelo para decidir pela aceitação ou não de um novo cliente. Para contornar esse problema, é necessário incluir no ajuste do modelo os clientes rejeitados. A dificuldade para isso é que não se conhece a performance desses clientes para classificá-los como bons/maus. Há alguns métodos para inferir qual seria o comportamento de um cliente recusado caso tivesse recebido o crédito solicitado. A esse procedimento denomina-se inferência de rejeitados. Esse procedimento não será detalhado nesse trabalho, pois o modelo desenvolvido destina-se somente à aplicação de clientes que contrataram o crédito.

2.4. TÉCNICAS ESTATÍSTICAS

A regressão logística (descrita adiante, no item 2.4.2) é a técnica mais utilizada nas empresas para desenvolvimento de modelos de *scoring* ((ROSA, 2000); (OHTOSHI, 2003)). Essa metodologia apresenta vantagens em relação a outros modelos de análise discriminante. Uma delas é que não se pressupõe que os dados de entrada tenham distribuição normal, embora seja desejável que tenham (HAIR *at al*, 2005). Além da regressão logística, há exemplos de aplicação de algoritmos genéticos (GONÇALVES, 2005), redes neurais ((GONÇALVES, 2005); (OHTOSHI, 2003)) e árvores de decisão (OHTOSHI, 2003), para avaliar risco de crédito, mas com foco em *credit score*, voltado para instituições financeiras.

2.4.1. Análise de *Cluster*

A Análise de *Cluster* é um conjunto de técnicas utilizado na identificação de padrões de comportamento em bancos de dados por meio da formação de grupos homogêneos (BARROSO; ARTES, 2003).

O objetivo da Análise de Cluster é agrupar as observações semelhantes de forma que os grupos sejam mais homogêneos internamente e diferentes entre si.

Sharma (1996), enumera os passos para a Análise de *Cluster* da seguinte forma:

1. Definição da medida de similaridade ou dissimilaridade entre os objetos.
2. Decisão de qual técnica de clusterização será utilizada: hierárquica ou não hierárquica.
3. Definição sobre o método de clusterização (ligação) para a técnica que já foi selecionada (exemplo: método do centroide para a técnica hierárquica de clusterização).
4. Decisão sobre o número de *clusters* a ser feito.
5. Interpretação da solução final.

2.4.1.1. Medidas de similaridade e dissimilaridade

Cada objeto pode ser representado por um ponto no espaço n-dimensional e, portanto, pode ser agrupado com outros que estejam próximos e que mais se assemelham a ele. Em algumas ocasiões, o interesse está no agrupamento de variáveis e considera-se estas no espaço p-dimensional.

Há dois tipos de medida usados para identificar pontos semelhantes: medidas de similaridade (quanto maior o valor, maior a semelhança entre os objetos) e medidas de dissimilaridade (quanto maior o valor, mais diferentes são os objetos) (Barroso & Artes, 2003). Frequentemente há muitas subjetividades na escolha das medidas de similaridade. Johnson & Wichern (1998) colocam importantes considerações para a escolha, incluindo a natureza das variáveis (discreta, contínua, binária) ou escalas de medida (nominal, ordinal, intervalar, razão) e conhecimento do assunto.

Quando objetos (unidades ou itens) são agrupados, a proximidade é normalmente indicada por algum tipo de distância. Por outro lado, variáveis são usualmente agrupadas com base nos coeficientes de correlação ou como medidas de associação (JOHNSON; WICHERN, 1998).

Na sequência são abordadas as distâncias em função do tipo de variável.

I. Variáveis quantitativas

As distâncias são as medidas de dissimilaridade mais utilizadas no estudo de bancos de dados compostos por observações de variáveis quantitativas. (BARROSO; ARTES, 2003)

Uma medida d_{ik} representa uma distância entre os pontos i e k se possuir as seguintes propriedades:

- a) a distância entre dois pontos é não negativa, ou seja, $d_{ik} \geq 0$ para qualquer escolha de i e k ;
- b) a distância de um ponto em relação a ele mesmo é zero, ou seja, $d_{ii} = 0$;
- c) a distância do ponto i até o ponto k é igual à distancia do ponto k até o ponto i , ou seja, $d_{ik} = d_{ki}$;

- d) a distância entre dois pontos é menor ou igual à soma da distância de cada um desses dois até um terceiro ponto, ou seja, $d_{ik} \leq d_{im} + d_{mk}$.

Algumas distâncias utilizadas são:

1) Distância euclidiana

A ideia básica é considerar cada observação como sendo um ponto no espaço euclidiano e, desse modo, a expressão determina a distância física entre os pontos. (Barroso & Artes, 2003). O problema dessa métrica é que ela não considera as variabilidades distintas das várias características. Assim, surgiu a distância estatística, descrita adiante. A expressão da distância euclidiana é:

$$d_{ik} = \sqrt{(\underline{x}_i - \underline{x}_k)^T (\underline{x}_i - \underline{x}_k)} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{kj})^2} \quad (7)$$

2) Distância estatística

Conforme descrito em Chaves Neto (2007), a distância estatística entre duas observações multivariadas é da forma:

$$d_{ik} = \sqrt{(\underline{x}_i - \underline{x}_k)' A^{-1} (\underline{x}_i - \underline{x}_k)} \quad (8)$$

para a qual A^{-1} é tal que $d(x_i, x_k) \geq 0$. Assim, $(\underline{x}_i - \underline{x}_k)' A^{-1} (\underline{x}_i - \underline{x}_k)$ é forma quadrática e as entradas A^{-1} , variâncias e covariâncias amostrais. Contudo, sem conhecimento dos grupos distintos, essas quantidades não podem ser calculadas, e então a distância euclidiana é preferível na Análise de *Cluster*.

3) Distância de Manhattan (*city block*)

A distância de Manhattan, também chamada de *city block*, é inspirada nos quarteirões de uma cidade, na qual os quarteirões são quadrados.

$$d_{ik} = \sum_{j=1}^p |x_{ij} - x_{kj}| \quad (9)$$

II. Variáveis categorizadas

O tratamento básico das variáveis qualitativas consiste na codificação de suas respostas por meio de variáveis indicadoras (*dummies*), ou seja, é designado o valor 0 ou 1 para indicar a presença ou ausência de certo atributo. Geralmente 1 indica a presença e 0 indica a ausência.

2.4.1.2. Métodos aglomerativos hierárquicos

Nesses métodos, os agrupamentos são formados a partir de uma matriz de distâncias ou de similaridade. Num primeiro passo, constrói-se a matriz de distâncias e procura-se nela o menor valor de distância — identifica-se assim o par de itens que será o mais semelhante em suas características. Então, os dois itens são agrupados e considerados como sendo um único objeto (ou grupo). Isso requer que se defina nova matriz de distância, eliminando-se os itens agrupados e incorporando-se o novo grupo (de dois). Em seguida, identifica-se o par mais semelhante, que formará novo grupo, e assim sucessivamente, até que todos os objetos estejam reunidos num mesmo grupo. Por meio da análise do histórico do agrupamento, pode-se definir *a posteriori* o número de grupos existentes nos dados. Maiores detalhes sobre esses métodos podem ser vistos em Barroso; Artes (2003). Uma técnica muito usada para se definir o número de *clusters* é passar uma reta pela metade da distância máxima no dendrograma.

O que diferencia os métodos de agrupamento são as regras para a ligação entre os grupos (itens). Adiante, são exemplificados alguns deles.

- 1) Método do vizinho mais próximo: a distância considerada é a menor distância entre um elemento de G1 (grupo 1) e um elemento de G2 (grupo 2), ou seja:
 $d(G1, G2) = \min (d_{ik})$, com $i \in G1$ e $k \in G2$.
- 2) Método do vizinho mais longe: define-se a distância como a maior entre um elemento de G1 e um elemento de G2, ou seja:
 $d(G1, G2) = \max (d_{ik})$, com $i \in G1$ e $k \in G2$
- 3) Método das médias das distâncias: nesse método, calcula-se a média das distâncias entre os elementos de G_1 e os de G_2 .

$$d(G_1, G_2) = \sum_{i \in G_1} \sum_{k \in G_2} \frac{d_{ik}}{g_1 g_2} \quad (10)$$

- 4) Método do centroide: esse método define a coordenada de cada grupo como sendo a média das coordenadas de seus objetos. Uma vez obtida essa coordenada, denominada centroide, a distância entre os grupos é obtida pelo cálculo das distâncias entre os centroides.
- 5) Método de Ward: a alocação de um elemento a um grupo é feita de modo a minimizar uma medida de homogeneidade interna. A cada etapa do método de Ward, busca-se agrupar objetos que tornem os agrupamentos formados homogêneos, tanto quanto possível. A medida de homogeneidade utilizada baseia-se na partição da soma de quadrados total de uma análise de variância. Como ilustração, Barroso; Artes (2008) consideram apenas a primeira variável do vetor de observações (X_1) e admitem a formação de k grupos. Nesse caso, a partição da soma de quadrados total é dada por:

$$SQT(1) = SQE(1) + SQD(1) \quad (11)$$

$$\sum_{j=1}^k \sum_{i \in G_j} (X_{i1} - \bar{X}_1)^2 = \sum_{j=1}^k n_j (X_{j1} - \bar{X}_1)^2 + \sum_{j=1}^k \sum_{i \in G_j} (X_{i1} - \bar{X}_{j1})^2 \quad (12)$$

em que $SQT(1)$ denota a soma de quadrados total da variável 1, $SQE(1)$ a soma de quadrado entre grupos da variável 1, $SQD(1)$ a soma de quadrados dentro do grupo da variável 1, G_j é o conjunto que indica os elementos do grupo j , n_j é o número de elementos do grupo j , \bar{X}_1 é a média da variável X_1 e \bar{X}_{j1} , a média da variável 1 no grupo j . Nessa partição, $SQD(1)$ mede o grau de homogeneidade interna dos grupos em relação a X_1 , enquanto $SQE(1)$ mede o grau de homogeneidade entre os grupos. Desse modo, uma boa partição para X_1 seria aquela que minimizasse $SQD(1)$ e, conseqüentemente, maximizasse $SQE(1)$.

Para considerar todas as variáveis simultaneamente, define-se a soma de quadrados da partição como:

$$SQDP = \sum_{i=1}^p SQD(i) \quad (13)$$

O primeiro passo do procedimento consiste na construção de $n-1$ grupos, sendo n o número de observações. Os próximos passos consistem na formação de $n-2$, $n-3$, ..., 1 grupos, sendo que o critério de seleção é a escolha do agrupamento com a menor SQDP em cada passo.

2.4.1.3. Comparação dos métodos hierárquicos

O método do vizinho mais longe tende a formar grupos mais homogêneos que o método do vizinho mais próximo. Isso se deve ao fato de ser um critério bastante rigoroso, pois a distância pequena entre dois grupos implica na proximidade de todos os elementos desse grupo.

O método de Ward é atraente por basear-se numa medida com forte apelo estatístico e por gerar grupos que, assim como os do método do vizinho mais longe, possuem alta homogeneidade interna. Os métodos hierárquicos produzem um gráfico muito útil na visualização dos grupos, chamado dendrograma.

2.4.1.4. Métodos não hierárquicos ou de partição

Técnicas não hierárquicas são utilizadas para formar k clusters de itens ou objetos.

1) Método das k-médias

Esse método exige que se fixe *a priori* o número de grupos que deve ser formado.

Numa versão simples, o processo é composto por três passos:

- a) Partição dos itens em k grupos iniciais;
- b) Prosseguimento com a lista de itens, colocando cada item no grupo cuja média (centroide) está mais próxima (usualmente calcula-se a distância euclidiana com observações padronizadas ou não). O centroide é recalculado para o grupo que recebeu um novo item e para o grupo que perdeu esse item.
- c) Repetição do segundo passo até que não restem realocações a serem feitas.

2) Método das k-medoides

O método das k-medoides é um método de partição baseado numa matriz de distâncias entre objetos. A medoide de um grupo é definida como o membro do grupo que possui a menor distância euclidiana média em relação aos demais membros do grupo. O critério de qualidade utilizado nesse método consiste na minimização da soma das distâncias entre as observações e as respectivas medoides.

Sendo k o número de grupos a ser formado, o algoritmo busca identificar k vetores observados que sejam representativos dos grupos (medoides). Desse modo, o critério de qualidade do método é dado por:

$$C = \sum_{i=1}^n C_i C_j = \min_{1 \leq i \leq k} d[m_i, j] \quad (14)$$

em que C é o critério de qualidade e $d[m_i, j]$ representa a distância entre a medoide $i(m_i)$ e a observação j. Uma vez identificados esses pontos, aloca-se cada objeto ao grupo de medoide mais próxima.

2.4.2. Regressão Logística

De acordo com Paula (2010), a regressão logística tem se constituído num dos principais métodos de modelagem estatística de dados quando a variável é dicotômica. Mesmo quando a resposta de interesse não é originalmente do tipo binário, alguns pesquisadores têm dicotomizado a resposta, de modo que a probabilidade de sucesso possa ser modelada pela regressão logística. Tudo isso se deve, principalmente, à facilidade de interpretação dos parâmetros de um modelo logístico e à possibilidade do uso desse tipo de metodologia em problemas de classificação. A questão de interpretação dos parâmetros é crucial, pois muitos modelos ou métodos utilizados no problema da classificação são de difícil interpretação.

A regressão logística é frequentemente apropriada para a análise de experimentos que apresentam variáveis respostas categóricas em que o interesse seja o de descrever a relação entre a variável resposta e um conjunto de variáveis explanatórias (GIOLO, 2006). Quando a variável resposta é dicotômica (somente duas

categorias), tem-se a denominada regressão logística dicotômica. Para variável resposta com mais de duas categorias, a denominação usada é regressão logística politômica. As variáveis explicativas, em regressão logística, podem ser categóricas ou contínuas. Variáveis *dummies* são usadas para que as variáveis explicativas categóricas possam ser consideradas no modelo de regressão logística.

No contexto apresentado em Chaves Neto (2007), o modelo linear logístico simples (uma variável explicativa) é derivado da função matemática conhecida como sigmoide:

$$f(y) = \frac{1}{1+e^{-y}}, y \in \mathbb{R}, \quad (15)$$

que varia monotonicamente de 0 a 1, à medida que y cresce, sendo simétrica em torno de $y = 1/2$.

Portanto:

$$f(y) = \frac{1}{1+e^{-y}} = \frac{e^y}{1+e^y} \quad (16)$$

e ainda a transformação LOGIT

$$f(y) = \ln \left[\frac{f(y)}{(1-f(y))} \right] = \ln \left[\frac{(1+e^{-y})^{-1}}{1-(1+e^{-y})^{-1}} \right] \quad (17)$$

$$\text{LOGIT } f(y) = \ln \left[\frac{1}{1+e^{-y}} / \left(1 - \frac{1}{1+e^{-y}}\right) \right] = \ln \left[\frac{1}{1+e^{-y}} / \frac{1+e^{-y}-1}{1+e^{-y}} \right] \quad (18)$$

$$= \ln \left[\frac{1}{1+e^{-y}} / \frac{e^{-y}}{1+e^{-y}} \right] = -\ln(1+e^{-y}) - (-y) - (-\ln(1+e^{-y})) \quad (19)$$

$$= -\ln(1+e^{-y}) + y + \ln(1+e^{-y}) = y \quad (20)$$

Então, impondo um modelo de regressão logística linear para estimar $P(Y = 1) = p(x)$, tratando o caso linear simples somente com uma variável explicativa, tem-se o modelo dado por:

$$\text{LOGIT } p(x) = \mu = \beta_0 + \beta_1 x \text{ que é o } y.$$

A aplicação desse modelo para $x = 0$ resulta:

$$p(0) = P(Y = 1 | x = 0) = \frac{e^{\beta_0 + \beta_1 \cdot 0}}{1 + e^{\beta_0 + \beta_1 \cdot 0}} = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \text{ da forma } \frac{e^y}{1 + e^y}.$$

em que β_0 e β_1 são parâmetros.

2.4.2.1. Modelo logístico linear múltiplo

Conforme Chaves Neto (2007), quando o interesse está em estabelecer a relação entre a variável resposta Y e as diversas variáveis explicativas X_1, X_2, \dots, X_{p-1} , que podem representar fatores de interesse, o modelo logístico linear múltiplo tem a forma:

$$\text{LOGIT } p(\underline{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} = \mu$$

ou

$$p(\underline{x}) = p(x_1, x_2, \dots, x_{p-1}) = \frac{e^\mu}{1 + e^\mu} = \frac{1}{1 + e^{-\mu}}$$

em que $\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} = \underline{x}' \underline{\beta}$ e $\underline{x}' = [x_1, x_2, \dots, x_{p-1}]$.

A diferença importante entre o modelo de regressão logística e o modelo de regressão linear geral é que a variável resposta na regressão logística assume apenas os valores 0 e 1, e no modelo linear geral a resposta assume valores de forma contínua. Em qualquer problema de regressão, a quantidade modelada é o valor médio da variável resposta, dado os valores conhecidos das variáveis independentes. Essa quantidade é chamada média condicional e será expressa por $E(Y|\underline{x})$, em que Y denota a variável resposta e \underline{x} é composto pelos valores das variáveis independentes. Em regressão linear, tem-se $-\infty < E(Y|\underline{x}) < +\infty$ e, em regressão logística, devido à natureza da variável resposta, $0 \leq E(Y|\underline{x}) \leq 1$. Observa-se que a mudança em $E(Y|\underline{x})$ por unidade de \underline{x} fica progressivamente menor quando $E(Y|\underline{x})$ torna-se próxima de zero ou de um. A curva em forma de “S” lembra a distribuição acumulada de uma variável aleatória, o que motivou o uso da distribuição logística para fornecer um modelo para $E(Y|\underline{x})$ (GIOLO, 2006).

Conforme mencionado anteriormente, a função logística toma valores entre zero e um; assume o valor zero numa parte do domínio das variáveis explicativas, e o valor um em outra parte do domínio. A função logística cresce suavemente na parte intermediária, possuindo particular curva em forma de “S”. Outras funções de

distribuição possuem as características mencionadas, no entanto a função logística foi escolhida basicamente por duas razões (GIOLO, 2006):

- a) do ponto de vista matemático, é extremamente flexível e fácil de ser usada;
- b) conduz a interpretações simples.

2.4.3. Variável *dummy*

A variável *dummy* também é conhecida como variável fictícia, muda e/ou indicadora. O uso de variáveis *dummies* para representar regressoras qualitativas num modelo de regressão é, em geral, mais indicado que proceder à alocação de códigos quaisquer às categorias dessas variáveis. A dificuldade básica com a alocação de códigos é que eles definem uma métrica para os níveis da variável qualitativa, métrica que pode não ser razoável (GIOLO, 2006).

Uma variável qualitativa com K classes será representada por K-1 variáveis indicadoras, cada qual tomando os valores 0 e 1.

2.4.4. Razão de chances: *odds ratio*

A *odds ratio* (OR) é a medida que descreve a intensidade de associação entre fatores em uma tabela de contingência dois por dois e é calculada por (GIOLO, 2006):

$$OR = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad (21)$$

em que:

n_{11} é a quantidade de elementos que pertence à linha 1 e coluna 1;

n_{12} é a quantidade de elementos que pertence à linha 1 e coluna 2;

n_{21} é a quantidade de elementos que pertence à linha 2 e coluna 1;

n_{22} é a quantidade de elementos que pertence à linha 2 e coluna 2.

A tabela de contingência 2 X 2 é ilustrada no quadro a seguir.

Quadro 3 — Tabela de contingência 2 X 2

Categorias da variável	Categorias de resposta		Total
	R ₁	R ₂	
C ₁	n ₁₁	n ₁₂	n ₁₊
C ₂	n ₂₁	n ₂₂	n ₂₊
Total	n ₊₁	n ₊₂	n

Essa medida pode ser usada em qualquer tipo de estudo e varia de 0 a infinito.

Analisando:

- Quando OR é igual a um, não existe associação entre as variáveis;
- Se OR for maior que 1, o grupo 1 tem chance (*odds*) maior de apresentar a resposta que o grupo 2;
- Se OR for menor que 1, o grupo 1 tem chance menor de apresentar a resposta que o grupo 2.

O intervalo de confiança de OR é determinado tomando-se o logaritmo neperiano, ou seja, $f = \ln(OR)$, cuja variância estimada é:

$$V(f) = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right) \quad (22)$$

O intervalo de confiança para o logaritmo neperiano da OR é expresso pela seguinte fórmula:

$$\left(f - Z_{\alpha/2} * \sqrt{V(f)} \leq \ln(OR) \leq f + Z_{\alpha/2} * \sqrt{V(f)} \right) = 1 - \alpha \quad (23)$$

em que $Z_{\alpha/2}$ denota o percentil da distribuição normal padrão correspondente ao nível de confiança de $1 - \alpha$.

2.4.5. Estimadores de máxima verossimilhança

A estimação dos parâmetros do modelo logístico linear com p parâmetros $p(x) = \frac{1}{1 + e^{-\mu}}$, onde $\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1}$ é feita aplicando-se o método da

máxima verossimilhança. Um tratamento adequado para esse modelo com $p-1$ variáveis explicativas é o da forma matricial, ou seja, define-se:

1. o vetor de variáveis explicativas de dimensão p por $\underline{x}' = [1 \ x_1 \ x_2 \ x_3 \ \dots \ x_{p-1}]$;
2. o vetor de parâmetros de dimensão p por $\underline{\beta}' = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_{p-1}]$;
3. o modelo completo por $p(\underline{x}) = \frac{1}{1 + e^{-\underline{x}\underline{\beta}'}}$.

Para aplicação desse método, é necessário inicialmente construir a função de verossimilhança, a qual expressa a probabilidade de ocorrência dos dados observados como uma função dos parâmetros desconhecidos. Os estimadores de máxima verossimilhança dos parâmetros são os valores que maximizam essa função (GIOLO, 2006).

Para obter as estimativas dos parâmetros deve-se considerar a variável resposta Y , que é dicotômica, ou seja, assume apenas os valores 1 e 0. Então, supõe-se que Y foi observada n vezes em correspondência a valores amostrados do vetor $\underline{x}' = [1 \ x_1 \ x_2 \ x_3 \ \dots \ x_{p-1}]$. Desta forma tem-se o vetor de respostas de dimensão n e a matriz do modelo de ordem $n \times p$. Considerando, ainda, que as n observações de Y estão repartidas por k grupos, ou seja, no ℓ -ésimo grupo representado por $\underline{x}'_{\ell} = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip-1}]$ são observados n_{ℓ} indivíduos, $\ell = 1, 2, \dots, k$. Deste modo se tem $\sum_{\ell=1}^k n_{\ell} = n$ e

$$\sum_{i=1}^{n_{\ell}} y_{i\ell} = y_{\ell}. \text{ Portanto, } y_{\ell} \sim b(n_{\ell}, p_{\ell}) \text{ com } p_{\ell} = p(\underline{x}_{\ell}).$$

Assumindo-se que as observações são independentes, e que β_j $j = 0, 2, \dots, p-1$ são os parâmetros a serem estimados, tem-se a seguinte expressão para a função de verossimilhança:

$$L(\underline{\beta}) = \sum_{\ell=1}^k \sum_{i=1}^{n_{\ell}} y_{\ell} x_{i\ell} \beta_i - \left[\sum_{\ell=1}^k n_{\ell} \ln(1 + e^{\sum_{i=1}^{p-1} x_{i\ell} \beta_i}) \right] \quad (24)$$

Prosseguindo com o método calcula-se as derivadas $\frac{\partial L(\underline{\beta})}{\partial \underline{\beta}} = D(\underline{\beta})$ e para achar o

ponto de inflexão faz-se $D(\hat{\underline{\beta}}) = 0$. E, então, deve-se resolver o sistema de equações:

$$X'(\underline{Y} - \underline{\mu}) = 0$$

onde $\underline{\mu}' = [\mu_1 \mu_2 \dots \mu_k]$, $\underline{Y}' = [Y_1 Y_2 \dots Y_k]$ e ainda $\mu_\ell = n_\ell p_\ell$. Consequentemente, dado a expressão de p_ℓ , o sistema deve ser resolvido por processo iterativo. Geralmente, aplica-se o Método de Newton-Raphson, porém pode-se aplicar o Algoritmo de Marquardt e, ainda, o Método do Score de Fisher com sucesso.

As estimativas de $\underline{\beta}$ são os valores que maximizam a função de verossimilhança.

O estimador de máxima verossimilhança tem muitas propriedades assintóticas, ou seja, da teoria das grandes amostras. Assim, o estimador é assintoticamente: consistente, não viciado e eficiente. Isso significa que, quanto maior o tamanho da amostra, mais próximo estará a estimativa do verdadeiro valor do parâmetro, a esperança do estimador é igual ao parâmetro, $E(\hat{\underline{\beta}}) = \underline{\beta}$ e é maior a precisão da estimativa. Ainda, esses estimadores EMV têm assintoticamente distribuição Gaussiana.

É necessário um tamanho grande de amostra para satisfazer essas propriedades, geralmente mais de 50 observações. Porém, existem trabalhos com $30 \leq n \leq 50$. Na existência de poucos dados, o método pode resultar em estimativas viciadas. Essa tendência pode causar discrepância na análise.

2.4.6. Seleção passo a passo: *stepwise*

Devido ao procedimento que avalia todas as regressões possíveis ser, em algumas situações, muito árduo, vários outros métodos têm sido desenvolvidos para um pequeno número de modelos de regressão, seja pela adição ou pela retirada de variáveis explicativas em cada tempo. Esses métodos são referidos como “procedimentos do tipo *stepwise*” (GIOLO, 2006).

O método baseia-se num algoritmo misto de inclusão e eliminação de covariáveis, segundo a importância destas e de acordo com algum critério estatístico. Esse grau de importância pode ser avaliado, por exemplo, pelo nível de significância do teste da razão de verossimilhança entre os modelos que incluem ou excluem as

covariáveis em questão. Quanto menor for esse nível de significância, tanto mais importante será considerada a covariável (PAULA, 2010). O método pode ser classificado em três categorias:

a) *Forward* — passo à frente

Esse procedimento começa sem nenhuma variável regressora no modelo de regressão. Um nível de significância α é, então, estabelecido a fim de que somente regressoras que apresentem significância menor ou igual a α possam fazer parte do modelo. Desse modo, têm-se os seguintes passos:

Passo 1: a primeira variável regressora a entrar no modelo é a que apresentar maior correlação simples com a resposta Y , isto é, maior coeficiente de correlação r_{YX_j} ($j = 1, \dots, k$) ou, equivalentemente, o menor p-valor associado ao teste F , tal que p-valor $\leq \alpha$ pré-estabelecido,

Passo 2: a segunda variável regressora a entrar no modelo é, agora, a com maior correlação parcial com a resposta Y , isto é, maior $r_{Y_j \cdot i}$ para $j \neq i$ e i o índice da regressora escolhida no primeiro passo. Em outras palavras, a segunda regressora escolhida é aquela com maior estatística F parcial, tal que p-valor $\leq \alpha$ pré-estabelecido.

Passos subsequentes: análogo ao segundo passo, as demais regressoras são inseridas uma a uma no modelo até que, num particular passo, todas as regressoras que ainda não estiverem no modelo apresentem p-valores associados aos testes F parciais $> \alpha$.

Note que, nesse procedimento, uma regressora escolhida em determinado passo para fazer parte do modelo permanece no modelo até o passo final, não havendo a possibilidade de exclusão desta em qualquer outro passo subsequente (GILOLO, 2006).

b) *Backward* — passo atrás

A eliminação *backward* começa com todas as k variáveis regressoras no modelo. Um nível de significância α é, então, estabelecido, a fim de que somente

regressoras que apresentem significância maior que α possam ser removidas do modelo. Deste modo, têm-se:

Passo 1: para cada variável regressora, é calculado a estatística F parcial como se ela fosse a última regressora a entrar no modelo. A primeira regressora a ser eliminada do modelo é aquela que apresentar o menor valor para a estatística F parcial ou, equivalentemente, o maior p-valor associado a essa estatística, tal que p-valor $> \alpha$ pré-estabelecido.

Passo 2 e passos subsequentes: o modelo sem a regressora escolhida no primeiro passo é ajustado, e novas estatísticas F parciais são calculadas para esse novo modelo. O procedimento usado no primeiro passo é então repetido, até que todas as regressoras que permanecerem no modelo apresentem p-valores associados aos testes F parciais menores ou iguais a α .

Nesse procedimento, uma regressora escolhida em determinado passo para ser removida do modelo não tem a possibilidade de vir a ser incluída em qualquer outro passo subsequente (GIOLO, 2006).

c) *Stepwise* — passo a passo

Assim como no procedimento de seleção *forward*, esse procedimento começa sem nenhuma regressora no modelo de regressão. Dois níveis de significância, α_1 e α_2 , são, no entanto, estabelecidos. Um deles para ser usado como critério de inclusão e o outro como critério de exclusão de regressoras. Alguns analistas escolhem $\alpha_1 = \alpha_2$, mas isso não é necessário. Os passos usados nesse procedimento são os seguintes:

Passo 1: a primeira regressora a entrar no modelo é a que apresentar maior correlação simples com a resposta Y, isto é, maior r_{YX_j} ($j = 1, \dots, k$) ou, equivalentemente, o menor p-valor associado ao teste F, tal que p-valor $\leq \alpha_1$ pré-estabelecido.

Passo 2: a segunda regressora a entrar no modelo é a que apresentar maior correlação parcial com a resposta Y, isto é, maior $r_{Y \cdot i}$ para $j \neq i$ e i o índice da regressora escolhida no primeiro passo. Em outras palavras, a segunda regressora escolhida é aquela com maior estatística F parcial, tal que p-valor $\leq \alpha_1$ pré-estabelecido.

Passo 3 e passos subsequentes: todas as regressoras são reavaliadas por meio de suas respectivas estatísticas F parciais. Havendo regressoras no modelo,

segundo α_2 estabelecido, que se mostrem redundantes na presença das demais, remove-se a que apresentar menor significância para o modelo. Caso não existam regressoras a serem removidas, analisam-se as estatísticas F parciais das regressoras que não estão no modelo. Existindo entre elas regressoras com p-valores que sejam $\alpha \leq 1$, inclui-se a que apresentar menor p-valor. O procedimento para quando não mais houver, segundo os níveis de significância α_1 e α_2 estabelecidos, regressoras que possam ser incluídas ou excluídas do modelo de regressão.

Nesse procedimento, uma regressora escolhida em determinado passo para fazer parte do modelo pode ser removida em outro passo. Em cada passo, pode ocorrer inclusão ou exclusão de determinada regressora, mas nunca a inclusão de uma e a exclusão de outra simultaneamente num mesmo passo (Giolo, 2006).

Os procedimentos de seleção *forward*, eliminação *backward* e seleção *stepwise* não necessariamente levam à escolha do mesmo modelo final.

Recomenda-se que todos os procedimentos sejam aplicados na esperança de que haja alguma concordância entre eles, ou mesmo para aprender algo a mais sobre a estrutura dos dados (GIOLO, 2006).

O procedimento de seleção *forward* tende a concordar com o de todas as regressões possíveis para subconjuntos pequenos de regressoras, enquanto o procedimento de eliminação *backward* para subconjuntos grandes de regressoras (GIOLO, 2006).

O modelo final obtido por qualquer um dos procedimentos deve ser analisado quanto ao seu sentido prático. Analistas inexperientes podem concluir por um modelo não realístico e sem aplicabilidade prática.

2.4.7. Significância das covariáveis no modelo

Após a obtenção das estimativas dos coeficientes β_j ($j = 0, 1, \dots, p-1$) avalia-se a adequação do modelo ajustado. Isto é feito pela Análise do Desvio (*Deviance Analysis*), que se constitui numa sequência de modelos encaixados e na verificação dos termos acrescentados. Assim, considerando dois modelos encaixados Mod_{p_1} e Mod_{p_2} sendo que Mod_{p_1} tem matriz X_{p_1} , contendo os parâmetros $\beta_1, \beta_2, \dots, \beta_{p_1}$ e o desvio D_{p_1} e

o Mod_{p_2} com matriz X_{p_2} , contendo os mesmo parâmetros que Mod_{p_1} e outros adicionais, $\beta_{p_1+1}, \beta_{p_1+2}, \dots, \beta_{p_2}$ e o desvio D_{p_2} . Testa-se, então, a hipótese nula:

$$H_0: \beta_{p_1} = \dots = \beta_{p_2} = 0$$

A Análise do Desvio (deviance analysis) é conhecida como ANODEV e tendo-se a sequência de modelos encaixados $M_{p_1}, M_{p_2}, \dots, M_{p_k}$ com os números de parâmetros $p_1 < p_2 < \dots < p_k$ e desvios $D_{p_1} > D_{p_2} > \dots > D_{p_k}$ deve-se construir uma tabela para a Análise do Desvio (ANODEV), que é uma generalização da ANOVA. Adiante, tem-se a tabela 4 da ANODEV. Então, o modelo nulo contém apenas a média e o modelo completo tem desvio zero.

Tabela 4 — ANODEV — Deviance Analysis

Modelo	Desvio	G.L.	Desvio	G.L.	Termo
1	D_{p_1}	p_1			
			$D_{p_1} - D_{p_2}$	$p_2 - p_1$	X_1
1 + X1	D_{p_2}	p_2			
			$D_{p_2} - D_{p_3}$	$p_3 - p_2$	X_2/X_1
1 + X1 + X2	D_{p_3}	p_3			
			$D_{p_3} - D_{p_4}$	$p_4 - p_3$	$X_3/X_1, X_2$
.....					
completo	0	0			

Outro modo é fazer a comparação de Mod_{p_1} com Mod_{p_2} usando a razão de máxima verossimilhança, uma vez que se sabe que $D_{p_1} - D_{p_2}$ tem distribuição qui-quadrado assintótica com $p_2 - p_1$ graus de liberdade, ou seja,

$$D_{p_1} - D_{p_2} \sim \chi_{p_2-p_1}$$

Em outras palavras, a comparação pode ser feita utilizando-se o teste da razão de verossimilhanças, em que a função de verossimilhança do modelo sem as covariáveis L_{SC} é comparada com a função de verossimilhança do modelo com as covariáveis L_{CC} . Formalmente, o teste é expresso por (GIOLO, 2006):

$$RV = -2 * \ln \left[\frac{\text{verossimilhança do modelo sem as covariáveis}}{\text{verossimilhança do modelo com as covariáveis}} \right]$$

O logaritmo da razão das verossimilhanças é multiplicada por -2 para que se obtenha uma quantidade cuja distribuição seja conhecida (no caso, a distribuição qui-quadrado), de modo que tal quantidade possa ser usada para a realização de testes de hipóteses. Em regressão logística, a estatística abaixo é chamada *deviance*.

$$D = -2 * \ln \left[\frac{\text{verossimilhança do modelo sob estudo}}{\text{verossimilhança do modelo saturado}} \right]$$

Um modelo saturado ou completo, é aquele que contém tantos parâmetros quantos os dados existentes. Assim, a estatística RV pode ser vista como a diferença de duas *deviances*. Sob a hipótese nula de que os p coeficientes associados às covariáveis no modelo são iguais a zero, a distribuição de RV será qui-quadrado com p graus de liberdade. Rejeição da hipótese nula, nesse caso, tem interpretação análoga àquela em regressão linear, ou seja, pode-se concluir que pelo menos um ou talvez todos os p coeficientes sejam diferentes de zero (Giolo, 2006).

A *deviance* (desvio) de um modelo compara o logaritmo da verossimilhança desse modelo com o logaritmo da verossimilhança do modelo saturado.

2.4.8. Diagnóstico do modelo ajustado

- a) Poder preditivo do modelo utiliza um ponto de corte a partir do qual se estabelece que:
 - i. a variável resposta recebe o valor 1, $Y=1$, para probabilidades estimadas pelo modelo maiores ou iguais a esse ponto de corte;
 - ii. a variável resposta receba o valor 0, $Y=0$, para probabilidades estimadas pelo modelo menores que esse ponto de corte (GIOLO, 2006).

b) Multicolinearidade

O Fator de Inflação de Variância (VIF) quantifica o grau em que cada variável é explicada pelas demais variáveis independentes. De forma simples, cada variável independente se torna uma dependente e passa por uma regressão em relação as demais variáveis independentes (Hair, Anderson, Tatham, & Black, 2005). Em Giolo, 2011, o VIF para o j -ésimo coeficiente de regressão é expresso por:

$$VIF_j = \frac{1}{1-R_j^2}, \quad (25)$$

em que R_j^2 é o coeficiente de determinação múltiplo obtido da regressão de X_j com as demais variáveis como regressoras. Assim, se X_j for quase linearmente dependente com alguma das outras regressoras, então R_j^2 será próximo de 1 e VIF_j será grande. Experiências práticas indicam que VIF maiores que 10 implicam que os coeficientes de regressão associados estão sendo pobremente estimados devido à multicolinearidade.

- c) Outras medidas auxiliares são obtidas por meio dos cálculos realizados para encontrar o poder preditivo, podendo-se encontrar (GIOLO, 2006):
- i. valor preditivo do modelo positivo;
 - ii. valor preditivo do modelo negativo;
 - iii. sensibilidade do modelo;
 - iv. especificidade do modelo.

A sensibilidade é definida como a proporção de resultados positivos que um teste apresenta quando realizado em sujeitos conhecidos que tenham a característica, ou seja, é a proporção de verdadeiros positivos.

A especificidade, por outro lado, é definida como a proporção de resultados negativos que um teste apresenta quando realizado em sujeitos conhecidos que estejam livres da característica (proporção de verdadeiros negativos).

O desejado num modelo é que ele tenha, simultaneamente, alta sensibilidade e especificidade.

2.4.9. Avaliação do erro do modelo ajustado — Método de Lachenbruch

Conforme descrito em Johnson; Wichern (1998), uma boa metodologia para estimar a taxa de erro do modelo é o método de validação de Lachenbruch. Esse método consiste em:

- a) manter fora da amostra o primeiro indivíduo e reajustar o modelo utilizando a partir do segundo até o último registro;
- b) aplicar o modelo reajustado no indivíduo que foi mantido fora da amostra;
- c) avaliar como o modelo classificou o indivíduo que estava fora da amostra;
- d) devolver o primeiro indivíduo que estava fora, retirar o segundo, reajustar o modelo, aplicar no segundo indivíduo que estava fora da amostra e avaliar a classificação;
- e) repetir o procedimento para cada observação da amostra.

Com os resultados da classificação pode-se formar uma matriz de confusão genérica conforme quadro a seguir:

Quadro 4 — Matriz de Classificação — Lachenbruch

Amostra observada	Amostra classificada pelo modelo		Total
	1	2	
1	n_{11}	n_{12}	N_1
2	n_{21}	n_{22}	N_2

sendo n_{ij} o número de elementos pertencentes à população de origem i e que são classificados pelo modelo como pertencentes à população j . Quando $i = j$, tem-se o número de classificações corretas, e quando $i \neq j$, tem-se o número de classificações incorretas.

3. MATERIAL E MÉTODOS

3.1. MATERIAL

Os dados disponíveis para o desenvolvimento da modelagem correspondem aos clientes que contrataram os serviços da prestadora no período entre janeiro e dezembro de 2007. No banco de dados disponível, foram incluídos os clientes que entraram na base nesse período, considerando-se apenas pessoa física (CPF). Esses clientes foram observados durante doze meses e, ao final, foram classificados como bons ou maus, de acordo com os pagamentos das faturas geradas.

3.1.1. Amostra

A amostra total é composta por 43.339 clientes que contrataram os serviços da credora entre janeiro e dezembro de 2007, distribuídos mensalmente conforme tabela 5.

Tabela 5 — Distribuição dos clientes por mês de contratação

Mês de contratação	Quantidade	Proporção
Jan-07	1.520	3,51%
Fev-07	1.593	3,68%
Mar-07	2.604	6,01%
Abr-07	2.139	4,94%
Mai-07	2.370	5,47%
Jun-07	2.619	6,04%
Jul-07	3.912	9,03%
Ago-07	5.435	12,54%
Set-07	4.020	9,28%
Out-07	5.796	13,37%
Nov-07	5.736	13,24%
Dez-07	5.595	12,91%
Total	43.339	100,00%

Fonte: a autora

Observa-se que os meses mais recentes na amostra têm maior representatividade. Isso se deve ao crescimento da empresa, que tende a ter mais contratações mensais.

3.1.2. Definição de bom e mau

A amostra dos clientes foi acompanhada por doze meses de performance para classificação de bom/mau. O critério utilizado foi:

- 1) Mau — atraso máximo no período de performance maior que 90 dias em pelo menos uma das faturas.
- 2) Bom — atraso máximo de até 5 dias nas faturas geradas no período de performance.
- 3) Indeterminado — clientes com atraso máximo entre 6 e 89 dias durante o período de performance.

Para definição desses critérios, foi considerado que a partir de 90 dias de atraso a empresa precisa alocar capital em PDD (Provisão para Devedores Duvidosos). Isso representa alto custo e reflete alta probabilidade de perda financeira.

Além disso, foi considerado que, a partir de cinco dias de atraso no pagamento, os clientes são enviados para a cobrança ativa, portanto geram custos.

Dessa forma, clientes com atraso maior que cinco dias não poderiam ser classificados como bons e foram incluídos na categoria indeterminados. Esse grupo não foi disponibilizado pela empresa, o que não gera impactos no ajuste do modelo, pois para isso são utilizados somente os clientes classificados como bons ou maus.

3.1.3. Variáveis

Para os clientes incluídos na amostra, foram consideradas as seguintes variáveis:

Quadro 5 — Variáveis analisadas

Variável	Descrição
CdCliente	Código do cliente
Resposta	Classificação do cliente bom/mau
Categoria	Categoria do cliente (residencial ou <i>retail</i>)

CEP	CEP de instalação do telefone
NAT_OCUP	Natureza da ocupação do cliente
OCUPACAO	Ocupação (profissão) do cliente
TpResidencia	Tipo de residência do cliente (própria, alugada, etc)
EstCivil	Estado civil do cliente
UF	Estado de instalação do telefone
Cidade	Cidade de instalação do telefone
CdProdVoz	Código do produto de voz
DsProdVoz	Descrição do produto de voz
TpProdVoz	Tipo do produto de voz
GpProdVoz	Grupo do produto de voz
SubGpProdVoz	Subgrupo do produto de voz
CdProdDados	Código do produto de dados
DsProdDados	Descrição do produto de dados
TpProdDados	Tipo do produto de dados
GpProdDados	Grupo do produto de dados
SubGpProdDados	Subgrupo do produto de dados
Parcelamento	Cliente com parcelamento (sim/não)
Internet	Cliente com internet (sim/não)
FlagFraude	Cliente que já teve algum indício de fraude (sim/não)
FlagDesc	Cliente que já fez algum pedido de desconexão (sim/não)
Migracao	Cliente que migrou de plano (sim/não)
StatusInst	Status do serviço prestado ao cliente (ativo, parcial, inativo)
TipoTecnologia	Tipo de tecnologia (cablo/antena)
CanalVendas	Canal de vendas
WhiteList	Lista com CPFs de clientes que, ao solicitar o serviço, já estavam pré-aprovados
Categoria	Categoria do serviço (residencial/retail)
VIFatura	Valor da fatura
VIAjuste	Valor do ajuste
TotalPago	Total pago pelo cliente
VIFaturaAj	Valor da fatura ajustado
MediaPago	Média dos valores pagos pelo cliente
MaiorVIPago	Maior valor de fatura já pago pelo cliente
QtdeLinhas	Quantidade de linhas telefônicas do cliente
VIFatAtrasoMax	Valor da fatura em que o cliente já teve o maior atraso
SeqFatura	Sequência da fatura (indica há quantos meses o cliente adquiriu o serviço)
L1feitos - L22 feitos	Quantidade de minutos de ligação feita por tipo de ligação
L1 recebidos - L22 recebidos	Quantidade de minutos de ligação recebida por tipo de ligação
Feitos_total	Total de ligações feitas (em minutos)
Recebidos_total	Total de ligações recebidas (em minutos)

Fonte: a autora

3.2. METODOLOGIA E PROCEDIMENTOS

Os modelos logísticos são ajustados a partir da amostra de clientes. As faturas maiores que zero geradas por esses clientes durante os doze meses de performance foram acompanhadas por 90 dias, para avaliar a performance de pagamento. Os clientes que pagaram todas as faturas com no máximo cinco dias de atraso foram classificados como bons, e os clientes que atingiram 90 dias ou mais de atraso em qualquer uma das faturas analisadas foram considerados maus. Os clientes com atraso entre 6 e 89 dias foram considerados indeterminados e não são utilizados para ajuste dos modelos.

Além da classificação de bom e mau, para cada cliente foram coletadas as variáveis cadastrais, comportamentais e de relacionamento com a empresa, descritas anteriormente.

Para identificar quantos grupos distintos de clientes existem na amostra, foi aplicada a técnica de Análise de *Cluster*. Cada grupo obtido na Análise de Cluster pode ser tratado como um segmento do modelo final ou pode-se incluir o resultado do *cluster* no modelo como uma variável preditora adicional.

Antes de ajustar o modelo final, as variáveis contínuas foram discretizadas. Todas as variáveis preditoras foram analisadas com o objetivo de fazer fusão de classes que possuam as mesmas características quanto à ocorrência de clientes bons e maus. Após a definição das classes para cada variável, são calculados os pesos de evidência de cada classe (WOE — *Weight of Evidence*) e criadas as variáveis *dummies*.

O ajuste do modelo final é realizado por meio da regressão logística por duas formas. A primeira considera como preditoras as variáveis *dummies*. Na segunda forma, são usadas como variáveis preditoras os pesos de evidência de cada classe (WOE — *Weight of Evidence*).

4. RESULTADOS E DISCUSSÃO

4.1. ANÁLISE DE *CLUSTER*

O objetivo da aplicação da Análise de Cluster foi agrupar clientes semelhantes e, desta forma, identificar e formar grupos homogêneos internamente e heterogêneos entre si.

A técnica que se revelou mais adequada, considerando a característica do problema, foi o Agrupamento Não-hierárquico. No entanto, para aplicar essa técnica foi necessário definir previamente o número de grupos e tentativamente observar os agrupamentos. Então, para visualizar um número de grupos razoável, foi aplicado Agrupamento Hierárquico utilizando a *procedure PROC CLUSTER* do software SAS. Dessa forma é possível construir o dendrograma. Foram consideradas três tipos de ligação para formar o agrupamento: método do centroide, método das médias das distâncias e o método de Ward que foi o preferido.

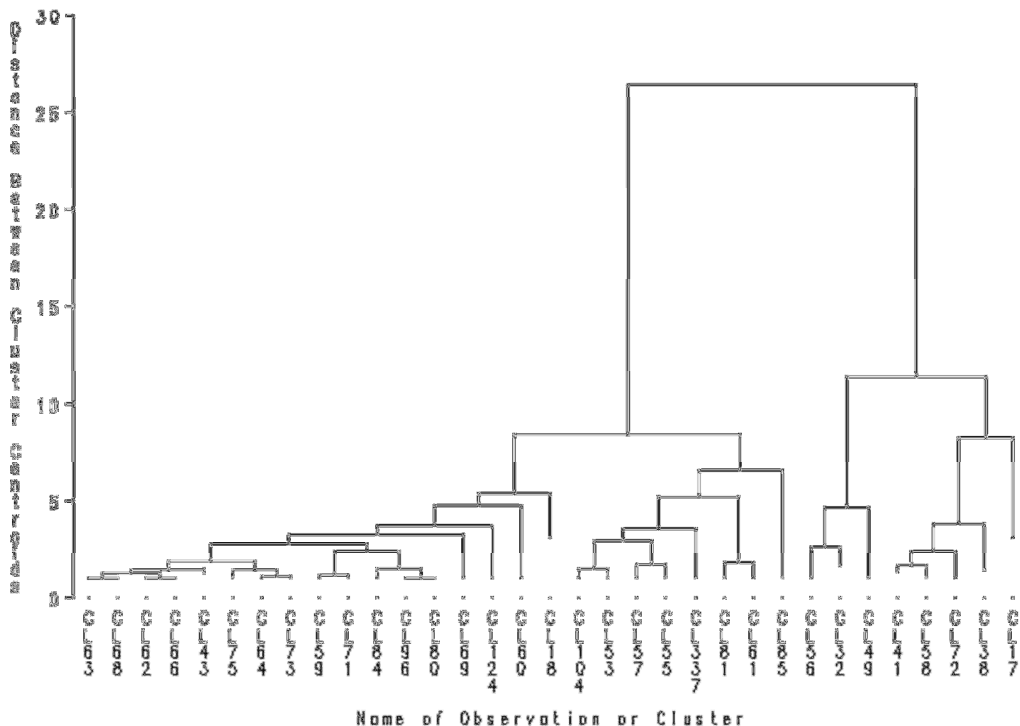


Figura 1 — Dendrograma — Método do centroide

De acordo com a figura 1, o grupo CL63 concentra 98,65% dos clientes, sendo assim, não é possível definir o número de *clusters* a ser formado utilizando o método do centroide, pois a maioria dos clientes concentra-se num único *cluster*.

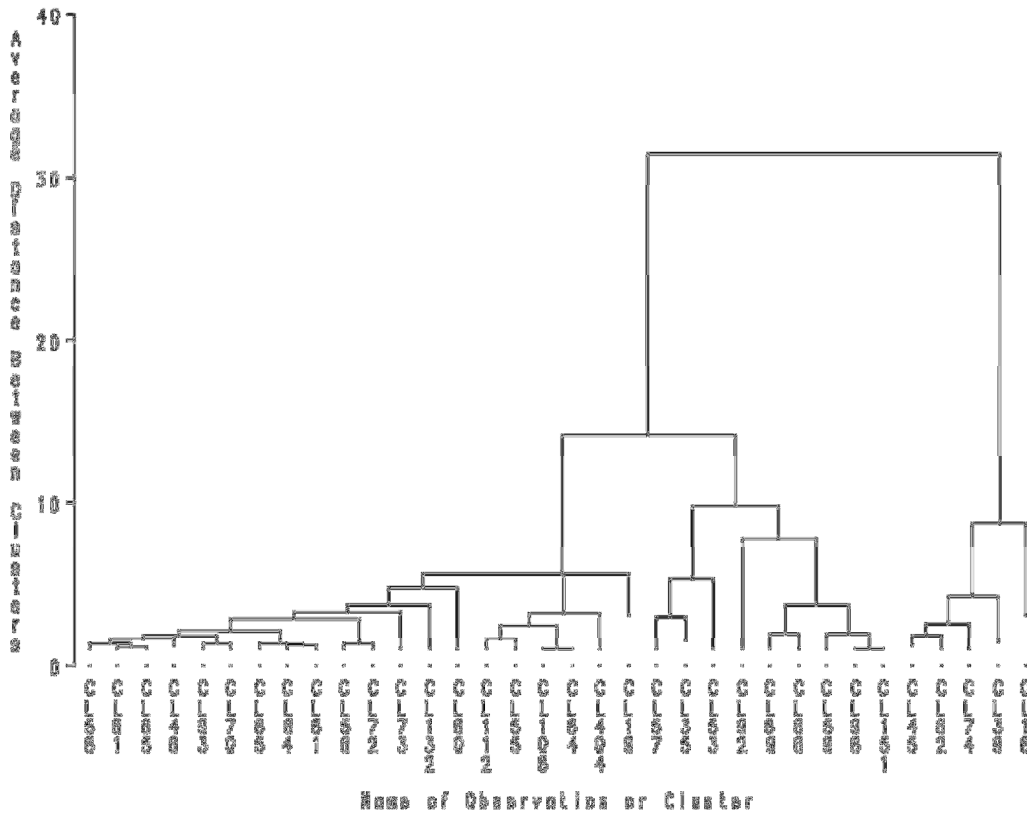


Figura 2 — Dendrograma — Método das médias das distâncias

De acordo com a Figura 2, o grupo CL58 concentra 99,10% dos clientes, sendo assim, não é possível definir o número de *clusters* a ser formado utilizando o método da média das distâncias, pois a maioria dos clientes concentra-se num único *cluster*.

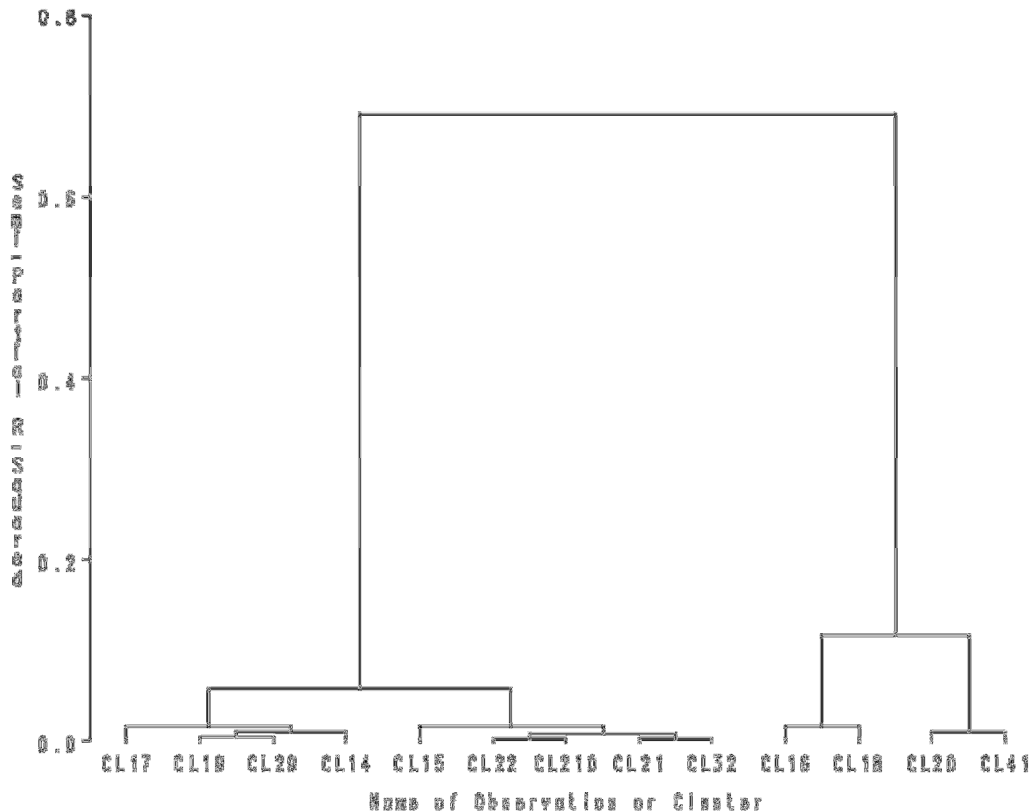


Figura 3 — Dendrograma — Método de Ward

De acordo com a Figura 3, o grupo CL17 concentra 94,62% dos clientes, sendo assim, não é possível definir o número de *clusters* a ser formado utilizando o método de Ward, pois a maioria dos clientes concentra-se num único *cluster*.

Mesmo não sendo possível definir o número adequado de grupos utilizando os três métodos hierárquicos, aplicou-se o método não hierárquico das k-médias para formação dos *clusters* através da *procedure PROC FASTCLUS* do software SAS. Como o número de grupos não ficou claro, foram testadas a criação de 2, 3, 4, ..., 50 *clusters*, utilizando o método k-médias. Observaram-se resultados semelhantes aos já encontrados por meio dos três métodos hierárquicos, sendo que, na maior quantidade de grupos testada, em que foram formados 50 *clusters*, ainda há concentração de 97,75% dos clientes num único grupo.

Com base nos resultados da Análise de Cluster, conclui-se que a amostra de clientes em estudo é toda um grupo homogêneo, ou seja, não se deve segmentá-lo.

4.2. TRATAMENTO DAS VARIÁVEIS CATEGÓRICAS

As variáveis categóricas que possuem várias classes precisam ser analisadas de modo que se possa fazer a fusão de algumas delas que têm proporção semelhante de clientes bons e, assim, reduzir o número de classes e, conseqüentemente, a quantidade de variáveis *dummies* a ser criada. Neste trabalho, para análise e fusão das classes das variáveis categóricas, foi construído inicialmente um intervalo de confiança para a proporção de clientes bons na amostra total.

Tabela 6 — Intervalo de confiança para a proporção de clientes bons

Dados	Valor
Amostra total	43.339
Bons	23.095
Maus	20.244
Proporção de clientes bons	53,2892%
Erro-padrão	0,2397%
Limite inferior do intervalo de confiança 95% para a proporção de clientes bons	52,8194%
Limite superior do intervalo de confiança 95% para a proporção de clientes bons	53,7589%
Range do intervalo de confiança 95% para a proporção de clientes bons	0,9395%

Fonte: a autora

Cada variável preditora foi analisada separadamente. Para cada classe (categoria) das variáveis preditoras, fez-se um intervalo de confiança para a proporção de clientes bons e adotados os seguintes critérios:

- a) Categorias com menos de 0,5% dos clientes foram classificadas como neutras.
- b) Categorias que têm entre 0,5% e 1% foram avaliadas se o seu intervalo de confiança cruza com intervalo de confiança geral da Tabela 6. Se cruzar com o intervalo geral, é classificada como categoria neutra. Se não cruzar com o intervalo geral, é classificada conforme Quadro 6.
- c) As categorias com 1% ou mais de clientes foram classificadas conforme o Quadro a seguir.

Quadro 6 — Critério para classificação das categorias das variáveis

Risco relativo	Classificação
< 0.50	07 – Horrível
≥ 0.50 e < 0.68	06 – Muito mau
≥ 0.68 e < 0.91	05 – Mau
≥ 0.91 e < 1.11	04 – Neutro
≥ 1.11 e < 1.51	03 – Bom
≥ 1.51 e ≤ 2.00	02 – Muito bom
> 2.00	01 – Excelente

Fonte: a autora

Para ilustrar esse procedimento, apresenta-se na Tabela 7 a classificação das categorias da variável estado (Unidade da Federação). Os números da classificação das categorias das demais variáveis estão nas tabelas do Apêndice 1.

Tabela 7 — Fusão das categorias da variável estado

Categoria	% Total	Risco relativo	Bom	Mau	% Bons	Intervalo de confiança 95%		Classificação
						Inferior	Superior	
AC – Acre	0.09	0.13	5	33	13%	2%	24%	04 – Neutro
DF – Distrito Federal	11.62	0.92	2.575	2.460	51%	50%	53%	04 – Neutro
GO – Goiás	8.57	0.68	1.623	2.089	44%	42%	45%	05 – Mau
MG – Minas Gerais	2.11	2.00	635	279	69%	66%	72%	02 – Muito bom
MS – Mato Grosso do Sul	4.47	0.89	976	960	50%	48%	53%	05 – Mau
MT – Mato Grosso	3.48	0.53	566	942	38%	35%	40%	06 – Muito mau
PR – Paraná	35.76	1.13	8.717	6.780	56%	55%	57%	03 – Bom
RO – Rondônia	0.36	0.18	26	130	17%	11%	23%	04 – Neutro
RS – Rio Grande do Sul	24.97	0.96	5.654	5.167	52%	51%	53%	04 – Neutro
SC – Santa Catarina	8.45	1.49	2.302	1.358	63%	61%	64%	03 – Bom
TO – Tocantins	0.13	0.17	9	46	16%	6%	26%	04 – Neutro

Fonte: a autora

Após aplicar os critérios de classificação, as categorias agrupadas e suas estatísticas são calculadas. Essa análise foi denominada análise bivariada e, para a variável estado, é ilustrada na Tabela 8. As tabelas com a análise bivariada para as demais variáveis estão no Apêndice 2.

Tabela 8 — Análise bivariada — Estado conforme os grupos de risco

Classificação	% Total	Maus	Bons	% Maus	% Bons	Taxa de maus	Odds	RR	WOE
00 – <i>Missing</i>	0%	0	1	0%	0%	0,00			
02 – Muito bom	2%	279	635	1%	3%	30,53	2,28	2,00	0,69
03 – Bom	44%	8.138	11.019	40%	48%	42,48	1,35	1,19	0,17
04 – Neutro	37%	7.836	8.275	39%	36%	48,64	1,06	0,93	-0,08
05 – Mau	13%	3.049	2.599	15%	11%	53,98	0,85	0,75	-0,29
06 – Muito mau	3%	942	566	5%	2%	62,47	0,60	0,53	-0,64

Fonte: a autora

Como ainda há categorias com baixas quantidades de clientes (menos de 5% do total), elas foram analisadas e reagrupadas conforme Tabela 9.

Tabela 9 — Reagrupamento das categorias da variável estado

Classificação	% Total	Maus	Bons	% Maus	% Bons	Taxa de maus	Odds	Risco relativo	WOE
02 – Muito bom /Bom	46%	8.417	11.654	42%	50%	41,94	1,38	1.21	0.19
04 – Neutro	37%	7.836	8.275	39%	36%	48,64	1,06	0.93	-0.08
05 – Mau/Muito Mau/ <i>missing</i>	17%	3.991	3.166	20%	14%	55,76	0,79	0.70	-0.36

Fonte: a autora

A tabela final com todos os agrupamentos para as variáveis está no Apêndice 3.

4.3. TRATAMENTO DAS VARIÁVEIS CONTÍNUAS (DISCRETIZAÇÃO)

A discretização das variáveis contínuas que compõem o modelo foi feita adotando-se como primeira premissa a criação de 20 faixas de valores que contêm 5% dos clientes em cada. Para isso, foi utilizada a *procedure* PROC RANK do *software* SAS. Para ilustrar esse procedimento, a Tabela 10 mostra o primeiro passo para a discretização da variável valor da fatura e os números obtidos. As estatísticas que compõem a Tabela 10 são: os valores máximo e mínimo das 43.339 faturas; os números de clientes maus e bons, em números absolutos e relativos; a *odds*; o WOE; o *information value* e o risco relativo com seu intervalo de confiança.

Tabela 10 — Discretização da variável valor da fatura

Mín.	Máx.	Maus	Bons	%Col Mau	%Col Bom	%Col Total	%Linha mau	%Linha bom	Odds	WOE	Inf. value	RR	RR IC low	RR IC Hight
0,00	20,62	648	1.518	3,2%	6,6%	5,0%	29,9%	70,1%	2,34	0,72	2,43	2,05	1,88	2,25
20,65	38,34	1.363	804	6,7%	3,5%	5,0%	62,9%	37,1%	0,59	-0,66	2,14	0,52	0,47	0,56
38,35	50,45	1.519	648	7,5%	2,8%	5,0%	70,1%	29,9%	0,43	-0,98	4,62	0,37	0,34	0,41
50,45	61,17	1.471	696	7,3%	3,0%	5,0%	67,9%	32,1%	0,47	-0,88	3,74	0,41	0,38	0,45
61,18	71,39	1.438	729	7,1%	3,2%	5,0%	66,4%	33,6%	0,51	-0,81	3,20	0,44	0,41	0,48
71,40	82,25	1.420	747	7,0%	3,2%	5,0%	65,5%	34,5%	0,53	-0,77	2,93	0,46	0,42	0,50
82,25	92,19	1.269	898	6,3%	3,9%	5,0%	58,6%	41,4%	0,71	-0,48	1,14	0,62	0,57	0,67
92,20	102,04	1.108	1.059	5,5%	4,6%	5,0%	51,1%	48,9%	0,96	-0,18	0,16	0,84	0,77	0,91
102,05	111,33	1.068	1.098	5,3%	4,8%	5,0%	49,3%	50,7%	1,03	-0,10	0,05	0,90	0,83	0,98
111,34	121,62	973	1.195	4,8%	5,2%	5,0%	44,9%	55,1%	1,23	0,07	0,03	1,08	0,99	1,17
121,63	132,38	971	1.196	4,8%	5,2%	5,0%	44,8%	55,2%	1,23	0,08	0,03	1,08	0,99	1,17
132,38	144,06	855	1.312	4,2%	5,7%	5,0%	39,5%	60,5%	1,53	0,30	0,43	1,35	1,24	1,46
144,06	156,77	884	1.283	4,4%	5,6%	5,0%	40,8%	59,2%	1,45	0,24	0,29	1,27	1,17	1,38
156,77	171,59	825	1.342	4,1%	5,8%	5,0%	38,1%	61,9%	1,63	0,35	0,62	1,43	1,31	1,55
171,60	188,24	739	1.428	3,7%	6,2%	5,0%	34,1%	65,9%	1,93	0,53	1,33	1,69	1,55	1,85
188,25	206,80	729	1.438	3,6%	6,2%	5,0%	33,6%	66,4%	1,97	0,55	1,44	1,73	1,58	1,89
206,82	231,57	643	1.524	3,2%	6,6%	5,0%	29,7%	70,3%	2,37	0,73	2,50	2,08	1,90	2,27
231,59	268,08	698	1.469	3,4%	6,4%	5,0%	32,2%	67,8%	2,10	0,61	1,78	1,84	1,69	2,01
268,10	334,53	684	1.483	3,4%	6,4%	5,0%	31,6%	68,4%	2,17	0,64	1,95	1,90	1,74	2,08
334,54	3.719,68	939	1.228	4,6%	5,3%	5,0%	43,3%	56,7%	1,31	0,14	0,09	1,15	1,06	1,25

Fonte: a autora

Após a criação das 20 faixas, nas quais cada uma contém 5% das faturas da amostra total de 43.339, são observados os intervalos de confiança para o risco relativo. Para as classes de valores vizinhas e seus respectivos intervalos de confiança que se cruzam, é feita a fusão das classes. Após essa fusão, a tabela é atualizada e suas estatísticas são recalculadas. Se houver mais algum intervalo de confiança para o risco relativo de faixas vizinhas que se cruzam, o procedimento é repetido. Repete-se a atualização até que cada faixa de valores possua intervalos de confiança para o risco relativo que não cruza com o intervalo de suas faixas vizinhas. Para a variável valor da fatura, a discretização final é dada na Tabela 11.

Tabela 11 — Discretização final — variável valor da fatura

Mín.	Máx.	Maus	Bons	%Col mau	%Col bom	%Col total	%Row mau	%Row bom	Odds	WOE	Inf. value	RR	RR IC low	RR IC High
0,00	20,62	648	1.518	3,2%	6,6%	5,0%	29,9%	70,1%	2,34	0,72	2,43	2,05	1,88	2,25
20,65	82,25	7.211	3.624	35,6%	15,7%	25,0%	66,6%	33,4%	0,50	-0,82	16,34	0,44	0,43	0,46
82,25	92,19	1.269	898	6,3%	3,9%	5,0%	58,6%	41,4%	0,71	-0,48	1,14	0,62	0,57	0,67
92,20	111,33	2.176	2.157	10,7%	9,3%	10,0%	50,2%	49,8%	0,99	-0,14	0,20	0,87	0,82	0,92
111,34	132,38	1.944	2.391	9,6%	10,4%	10,0%	44,8%	55,2%	1,23	0,08	0,06	1,08	1,02	1,14
132,38	171,59	2.564	3.937	12,7%	17,0%	15,0%	39,4%	60,6%	1,54	0,30	1,30	1,35	1,29	1,41
171,60	3.719,68	4.432	8.570	21,9%	37,1%	30,0%	34,1%	65,9%	1,93	0,53	8,03	1,69	1,64	1,75

Fonte: a autora

Para todas as variáveis contínuas, foi adotado esse mesmo procedimento para a discretização. A tabela com as estatísticas para as demais variáveis está no Apêndice 4.

Após a análise e o tratamento de cada uma das variáveis preditoras (contínuas e categóricas), foram criadas variáveis *dummies* para as categorias das variáveis. Além disso, para cada variável, foi criada na base de dados uma coluna com o WOE correspondente a cada categoria da variável.

4.4. AJUSTE DO MODELO USANDO VARIÁVEIS *DUMMIES*

Os ajustes dos modelos foram feitos utilizando a *procedure* PROC LOGISTIC do *software* SAS. Para seleção inicial das variáveis, aplicou-se o método *backward*. Após a seleção inicial, avaliou-se o VIF (Fator de Inflação de Variância), para avaliar a multicolinearidade no modelo final.

Para o modelo teórico definido como:

$$\theta(x) = P(Y = 1|\underline{x}) = \frac{e^{\beta_0 + \sum_{k=1}^{p-1} \beta_k}}{1 + e^{\beta_0 + \sum_{k=1}^{p-1} \beta_k}} = \frac{e^{\underline{\beta}' \underline{x}}}{1 + e^{\underline{\beta}' \underline{x}}} \quad (26)$$

os parâmetros estimados são apresentados no Quadro 7.

Quadro 7 — Modelo ajustado com variáveis *dummies* — com intercepto

Variável	Dummy	$\hat{\beta}$	p-valor	VIF	Odds ratio
Intercepto		0,03223	0,84258	0,00000	
Parcelamento	1) Sim	-7,73612	0,00000	1,58472	0,00044
FlagFraude	2) Sim	-1,25824	0,00000	1,34656	0,28415
FlagDesc	2) Sim	1,88957	0,00000	1,09680	6,61653
VIFatura	01. $\leq 20,62$	4,55680	0,00000	1,82104	95,27779
	02. $> 20,62$ e $\leq 82,25$	1,02559	0,00000	2,39804	2,78875
	06. $> 132,38$ e $\leq 171,59$	-0,85593	0,00000	1,90623	0,42489
	07. $> 171,59$	-2,96616	0,00000	4,91642	0,05150
VIAjuste	02. $> 1,49$ e $\leq 5,28$	0,35580	0,00252	1,02168	1,42733
TotalPago	01. $\leq 1,90$	-1,92001	0,00000	3,31096	0,14661
	02. $> 1,90$ e $\leq 40,40$	-1,56694	0,00000	2,31518	0,20868
	08. $> 141,30$ e $\leq 178,64$	2,59214	0,00000	1,81995	13,35839
	09. $> 178,64$	4,05094	0,00000	4,49039	57,45165
MediaPago	00. Missing	-2,21561	0,00000	2,13245	0,10909
	04. $> 83,19$ e $\leq 90,71$	0,28425	0,01253	1,21921	1,32876
	05. $> 90,71$ e $\leq 97,48$	0,68276	0,00000	1,25267	1,97933
	06. $> 97,48$ e $\leq 110,30$	0,89549	0,00000	1,57432	2,44854
	07. $> 110,30$ e $\leq 222,56$	1,11686	0,00000	2,72894	3,05523
MaiorVIPago	03. $> 47,21$ e $\leq 101,91$	0,67833	0,00000	2,25411	1,97058
	04. $> 101,91$ e $\leq 123,78$	0,69321	0,00000	1,77883	2,00013
	05. $> 123,78$ e $\leq 242,26$	0,29722	0,00035	2,33936	1,34611
QtdeLinhas	02. > 1	0,31094	0,00001	2,05185	1,36471
VIFatAtrasoMax	04. $> 214,84$ e $\leq 250,43$	-0,75881	0,00000	1,11223	0,46822
	05. $> 250,43$ e $\leq 313,86$	-0,50537	0,00048	1,19325	0,60328
	06. $> 313,86$ e $\leq 455,93$	-0,69156	0,00004	1,31505	0,50079
	07. $> 455,93$	-1,31253	0,00000	1,61469	0,26914
SeqFatura	01. ≤ 9	-1,75099	0,00000	3,69132	0,17360
	02. > 9 e ≤ 10	-0,60535	0,00000	1,38075	0,54588
	04. > 12 e ≤ 14	0,85334	0,00000	1,50939	2,34748
	05. > 14 e ≤ 17	2,06669	0,00000	1,74228	7,89862
	06. > 17	2,84545	0,00000	2,24635	17,20926
L1feitos	01. $\leq 96,55$	-0,33503	0,00000	2,11926	0,71531
	04. $> 236,58$	0,48028	0,00000	1,86060	1,61653
L2feitos	03. $> 44,80$ e $\leq 516,14$	-0,31510	0,00879	1,09300	0,72971
	04. $> 516,14$	-0,63158	0,00012	1,74595	0,53175
L7feitos	01. $\leq 3,37$	1,18663	0,00000	1,68430	3,27603
	02. $> 3,37$ e $\leq 6,69$	0,87880	0,00000	1,31036	2,40800
	03. $> 6,69$ e $\leq 10,91$	0,46557	0,00000	1,28243	1,59292
	06. $> 27,72$ e $\leq 33,52$	-0,38854	0,00072	1,15429	0,67805

	07. > 33,52 e ≤ 68,46	-0,66691	0,00000	1,49239	0,51329
	08. > 68,46 e ≤ 100,67	-0,72230	0,00000	1,29786	0,48564
	09. > 100,67 e ≤ 186,54	-0,94642	0,00000	1,46595	0,38813
	10. > 186,54	-0,96894	0,00000	2,25130	0,37949
L8feitos	03. > 3,56 e ≤ 9,85	-0,26731	0,00042	1,16257	0,76544
	04. > 9,85 e ≤ 15,83	-0,44354	0,00028	1,08780	0,64176
	05. > 15,83 e ≤ 31,89	-0,43026	0,00105	1,11862	0,65034
	06. > 31,89	-0,54058	0,00024	1,26785	0,58241
L9feitos	03. > 12,02 e ≤ 29,96	0,27808	0,00000	1,35653	1,32059
	04. > 29,96 e ≤ 52,86	0,36463	0,00001	1,32069	1,43998
	05. > 52,86	0,39762	0,00132	1,16116	1,48827
L10feitos	01. ≤ 2,78	0,27133	0,00005	1,84768	1,31171
	03. > 5,99	-0,15646	0,03020	1,81318	0,85517
L18feitos	03. > 1,59 e ≤ 7,94	0,18853	0,00484	1,05027	1,20747
L19feitos	01. ≤ 0,48	-0,19520	0,00058	1,28206	0,82267
L20feitos	03. > 1,68	0,35035	0,00000	1,40198	1,41956
L7recebidos	07. > 78,35	0,40939	0,00824	1,31361	1,50589
L20recebidos	03. > 0,87 e ≤ 2,16	-0,30476	0,00006	1,34803	0,73730
	04. > 2,16 e ≤ 7,35	-0,72654	0,00000	1,61005	0,48358
	05. > 7,35 e ≤ 12,98	-0,48865	0,00000	1,33735	0,61346
	06. > 12,98	-0,34077	0,00016	1,61175	0,71123
L21recebidos	04. > 20,02 e ≤ 58,59	0,21870	0,01623	1,04015	1,24446
Feitos_total	01. ≤ 144,94	-0,65459	0,00000	2,16454	0,51966
	03. > 270,72 e ≤ 732,55	0,32532	0,00005	2,44758	1,38447
	04. > 732,55	0,76726	0,00000	3,26410	2,15386
GROUPING - CanalVendas	BCENT	-0,14259	0,01164	1,32806	0,86711
	Outros	0,33001	0,00006	1,24031	1,39098
GROUPING - CL_CdProdVoz	01-Excelente	0,67921	0,00000	1,22817	1,97233
GROUPING - CL_CEP_C	06-Muito mau	-0,18836	0,02255	1,32683	0,82832
	07-Horrível/Missing	-0,29742	0,02772	1,71898	0,74274
GROUPING - CL_Cidade	01-Excelente/Muito bom	0,48781	0,00000	1,18518	1,62874
	05-Mau	0,20630	0,00155	1,65189	1,22912
	06-Muito mau/Horrível/Missing	0,35492	0,01426	2,16211	1,42607
GROUPING - CL_DsProdVoz	06-Muito mau/Horrível/Missing	-0,53676	0,00000	1,49373	0,58464
GROUPING - CL_NAT_OCUP	02-Muito bom/Missing	0,46010	0,00005	1,08288	1,58423
GROUPING - CL_OCUPACAO	01-Excelente/Muito bom	0,24907	0,01072	1,11588	1,28283
	03-Bom	0,28361	0,00063	1,10919	1,32792
	06-Muito mau/Mau/Horrível	-0,25761	0,00288	1,04904	0,77290
GROUPING - CL_UF	05-Mau/Muito mau/missing	-0,46145	0,00000	1,70392	0,63037
GROUPING - Migracao	1) Sim	0,37821	0,00000	1,13980	1,45967

GROUPING - WhiteList	01 - 0 a 2	1,57657	0,00000	1,26630	4,83832
	02 - 3 a 5	0,82863	0,00000	1,15282	2,29017

Fonte: a autora

Observa-se que o modelo ajustado apresenta ajuste satisfatório. Todos os parâmetros são significativos com valor-p $p < 0,05$, exceto o intercepto que foi aceito como nulo, dado o seu valor-p $p = 0,84258$, considerando nível de significância de $0,05$. Por esse motivo, o modelo foi reajustado sem o intercepto adiante no quadro 9. Obteve-se, também, o quadro da Análise do Desvio, a seguir.

Quadro 8 — Análise do Desvio - Modelo com variáveis *dummies* — com intercepto

Fonte de variação	<i>Deviance</i>	Graus de liberdade	p-valor
Modelo	48.379,1	82	0,0000
Resíduos	11.513,8	43.256	1,0000
Total	59.892,9	43.338	

Fonte: a autora

Como o valor-p para o modelo ajustado na Análise do Desvio é $p = 0,0000 < 0,05$, pode-se concluir que existe uma relação estatisticamente significativa entre a variável resposta Y e as variáveis explicativas que compõem o vetor \underline{x} . Além disso, o valor-p para os resíduos é $p = 1$, o que indica que o modelo não é significativamente pior do que o melhor modelo possível para esses dados. Com base nesses resultados, também é possível observar que a porcentagem do Desvio explicada pelo modelo é igual a 80,77%. Esta estatística é semelhante ao R^2 da regressão linear.

Quadro 9 — Modelo ajustado com variáveis *dummies* — sem intercepto

Variável	<i>Dummy</i>	$\hat{\beta}$	p-valor	VIF	<i>Odds ratio</i>
Parcelamento	1) Sim	-7,73781	0,00000	1,58	<0.001
FlagFraude	2) Sim	-1,25798	0,00000	1,35	0,28
FlagDesc	2) Sim	1,88968	0,00000	1,10	6,62
VIFatura	01. $\leq 20,62$	4,56403	0,00000	1,82	95,97
	02. $> 20,62$ e $\leq 82,25$	1,02997	0,00000	2,40	2,80
	06. $> 132,38$ e $\leq 171,59$	-0,85369	0,00000	1,91	0,43
	07. $> 171,59$	-2,96388	0,00000	4,92	0,05
VIAjuste	02. $> 1,49$ e $\leq 5,28$	0,35635	0,00248	1,02	1,43
TotalPago	01. $\leq 1,90$	-1,91493	0,00000	3,31	0,15

	02. > 1,90 e ≤ 40,40	-1,56662	0,00000	2,32	0,21
	08. > 141,30 e ≤ 178,64	2,59330	0,00000	1,82	13,37
	09. > 178,64	4,05412	0,00000	4,49	57,63
MediaPago	00. Missing	-2,21613	0,00000	2,13	0,11
	04. > 83,19 e ≤ 90,71	0,28772	0,01053	1,22	1,33
	05. > 90,71 e ≤ 97,48	0,68641	0,00000	1,25	1,99
	06. > 97,48 e ≤ 110,30	0,90004	0,00000	1,57	2,46
	07. > 110,30 e ≤ 222,56	1,12178	0,00000	2,73	3,07
MaiorVIPago	03. > 47,21 e ≤ 101,91	0,68695	0,00000	2,25	1,99
	04. > 101,91 e ≤ 123,78	0,70164	0,00000	1,78	2,02
	05. > 123,78 e ≤ 242,26	0,30355	0,00008	2,34	1,36
QtdeLinhas	02. > 1	0,31406	0,00000	2,05	1,37
VIFatAtrasoMax	04. > 214,84 e ≤ 250,43	-0,75780	0,00000	1,11	0,47
	05. > 250,43 e ≤ 313,86	-0,50326	0,00049	1,19	0,61
	06. > 313,86 e ≤ 455,93	-0,68891	0,00004	1,32	0,50
	07. > 455,93	-1,30949	0,00000	1,61	0,27
SeqFatura	01. ≤ 9	-1,74844	0,00000	3,69	0,17
	02. > 9 e ≤ 10	-0,60201	0,00000	1,38	0,55
	04. > 12 e ≤ 14	0,85615	0,00000	1,51	2,35
	05. > 14 e ≤ 17	2,06999	0,00000	1,74	7,93
	06. > 17	2,84824	0,00000	2,25	17,26
L1feitos	01. ≤ 96,55	-0,33308	0,00000	2,12	0,72
	04. > 236,58	0,48088	0,00000	1,86	1,62
L2feitos	03. > 44,80 e ≤ 516,14	-0,31485	0,00884	1,09	0,73
	04. > 516,14	-0,63125	0,00012	1,75	0,53
L7feitos	01. ≤ 3,37	1,18921	0,00000	1,68	3,28
	02. > 3,37 e ≤ 6,69	0,88123	0,00000	1,31	2,41
	03. > 6,69 e ≤ 10,91	0,46790	0,00000	1,28	1,60
	06. > 27,72 e ≤ 33,52	-0,38627	0,00073	1,15	0,68
	07. > 33,52 e ≤ 68,46	-0,66432	0,00000	1,49	0,52
	08. > 68,46 e ≤ 100,67	-0,71900	0,00000	1,30	0,49
	09. > 100,67 e ≤ 186,54	-0,94312	0,00000	1,47	0,39
	10. > 186,54	-0,96637	0,00000	2,25	0,38
L8feitos	03. > 3,56 e ≤ 9,85	-0,26586	0,00043	1,16	0,77
	04. > 9,85 e ≤ 15,83	-0,44218	0,00029	1,09	0,64
	05. > 15,83 e ≤ 31,89	-0,42863	0,00108	1,12	0,65
	06. > 31,89	-0,53862	0,00025	1,27	0,58
L9feitos	03. > 12,02 e ≤ 29,96	0,27949	0,00000	1,36	1,32
	04. > 29,96 e ≤ 52,86	0,36556	0,00001	1,32	1,44
	05. > 52,86	0,39817	0,00130	1,16	1,49
L10feitos	01. ≤ 2,78	0,27563	0,00001	1,85	1,32

	03. > 5,99	-0,15295	0,02882	1,81	0,86
L18feitos	03. > 1,59 e ≤ 7,94	0,18955	0,00449	1,05	1,21
L19feitos	01. ≤ 0,48	-0,19282	0,00050	1,28	0,83
L20feitos	03. > 1,68	0,35136	0,00000	1,40	1,42
L7recebidos	07. > 78,35	0,40968	0,00820	1,31	1,51
L20recebidos	03. > 0,87 e ≤ 2,16	-0,30252	0,00006	1,35	0,74
	04. > 2,16 e ≤ 7,35	-0,72417	0,00000	1,61	0,49
	05. > 7,35 e ≤ 12,98	-0,48685	0,00000	1,34	0,62
	06. > 12,98	-0,33920	0,00016	1,61	0,71
L21recebidos	04. > 20,02 e ≤ 58,59	0,21932	0,01586	1,04	1,25
Feitos_total	01. ≤ 144,94	-0,65273	0,00000	2,16	0,52
	03. > 270,72 e ≤ 732,55	0,32836	0,00003	2,45	1,39
	04. > 732,55	0,77005	0,00000	3,26	2,16
GROUPING - CanalVendas	BCENT	-0,13970	0,01051	1,33	0,87
	Outros	0,33241	0,00005	1,24	1,39
GROUPING - CL_CdProdVoz	01-Excelente	0,68068	0,00000	1,23	1,98
GROUPING - CL_CEP_C	06-Muito Mau	-0,18787	0,02284	1,33	0,83
	07-Horrível/Missing	-0,29646	0,02812	1,72	0,74
GROUPING - CL_Cidade	01-Excelente/Muito Bom	0,48966	0,00000	1,19	1,63
	05-Mau	0,20741	0,00140	1,65	1,23
	06-Muito Mau/Horrível/Missing	0,35518	0,01419	2,16	1,43
GROUPING - CL_DsProdVoz	06-Muito Mau/Horrível/Missing	-0,53438	0,00000	1,49	0,59
GROUPING - CL_NAT_OCUP	02-Muito Bom/Missing	0,46144	0,00004	1,08	1,59
GROUPING - CL_OCUPACAO	01-Excelente/Muito Bom	0,25005	0,01032	1,12	1,28
	03-Bom	0,28512	0,00056	1,11	1,33
	06-Muito Mau/Mau/Horrível	-0,25626	0,00295	1,05	0,77
GROUPING - CL_UF	05-Mau/Muito Mau/missing	-0,46168	0,00000	1,70	0,63
GROUPING - Migracao	1) Sim	0,37885	0,00000	1,14	1,46
GROUPING - WhiteList	01 - 0 a 2	1,57767	0,00000	1,27	4,84
	02 - 3 a 5	0,82984	0,00000	1,15	2,29

Fonte: a autora

No primeiro ajuste sem o intercepto, a variável " L2feitos : 02. > 0,00 and <= 44,80" resultou valor-p $p = 0,0788$, que não é significativo ao nível de 0,05 e foi retirada do modelo. Observa-se que o modelo ajustado apresenta ajuste satisfatório. Todos os parâmetros são significativos com valor-p $p < 0,05$, considerando nível de significância de 0,05. Obteve-se, também, o quadro da Análise do Desvio, a seguir.

Quadro 10 — Análise do Desvio - Modelo com variáveis *dummies* — sem intercepto

Fonte de variação	<i>Deviance</i>	Graus de liberdade	p-valor
Modelo	48.566,8	80	0,0000
Resíduos	11.513,8	43.258	1,0000
Total	60.080,6	43.338	

Fonte: a autora

Como o valor-p para o modelo ajustado na Análise do Desvio é $p = 0,0000 < 0,05$, pode-se concluir que existe uma relação estatisticamente significativa entre a variável resposta Y e as variáveis explicativas que compõem o vetor \underline{x} . Além disso, o valor-p para os resíduos é $p = 1$, o que indica que o modelo não é significativamente pior do que o melhor modelo possível para esses dados. Com base nesses resultados, também é possível observar que a porcentagem do Desvio explicada pelo modelo é igual a 80,83%. Esta estatística é semelhante ao R^2 da regressão linear.

Observando o modelo final ajustado com variáveis *dummies*, é possível identificar como principais fatores de risco de inadimplência: os parcelamentos, os alertas de fraude, os altos valores de fatura, os baixos valores de pagamentos históricos, os altos valores da fatura paga com o maior número de dias em atraso, o tempo de relacionamento (clientes novos representam maior risco), determinadas regiões identificadas por meio de estado, cidade e CEP, além do tipo de serviço contratado e do perfil de consumo.

Para visualizar os resultados do modelo quanto a classificação dos clientes, o gráfico que acumula a quantidade de clientes Bons e Maus de acordo com a probabilidade de pagamento é adequado. Observando o gráfico é possível visualizar que os clientes Maus estão concentrados nas baixas probabilidades de pagamento, enquanto que os clientes bons se concentram nas altas probabilidades de pagamento.

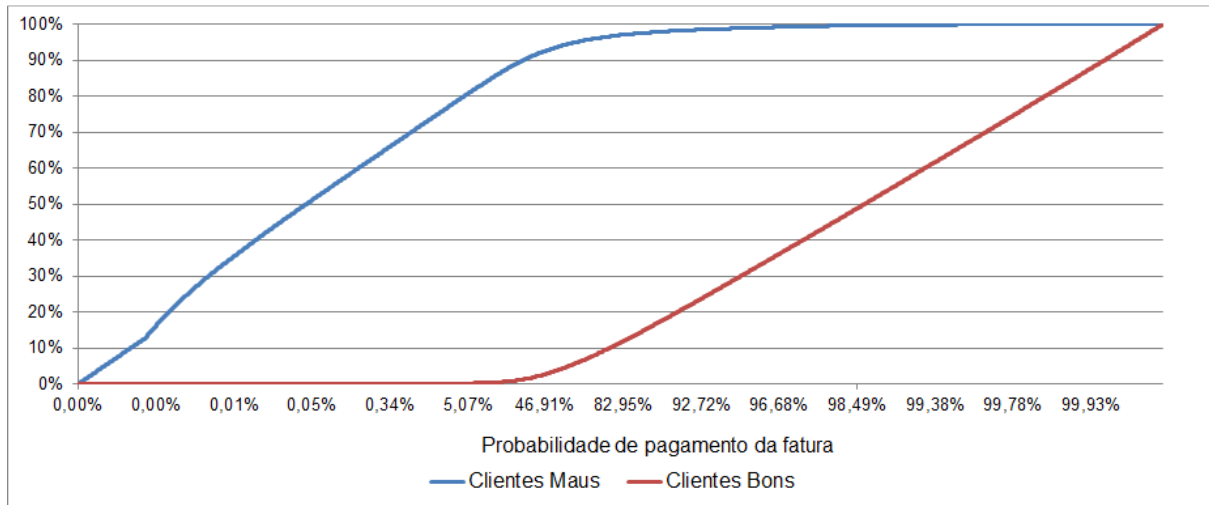


Figura 4 — Distribuição acumulada para os clientes Bons e Maus pela probabilidade de pagamento da fatura — Modelo com variáveis *dummies*

4.5. AJUSTE DO MODELO USANDO O WOE COMO VARIÁVEL PREDITORA

Para o ajuste dos modelos, foi utilizada a *procedure* PROC LOGISTIC do *software* SAS. Para seleção inicial das variáveis, utilizou-se o método *backward*. Após a seleção inicial, avaliou-se o VIF (Fator de Inflação de Variância) para avaliar multicolinearidade no modelo final.

Para o modelo teórico definido como,

$$\theta(x) = P(Y = 1|x) = \frac{e^{\beta_0 + \sum_{k=1}^{p-1} \beta_k}}{1 + e^{\beta_0 + \sum_{k=1}^{p-1} \beta_k}} = \frac{e^{\beta'x}}{1 + e^{\beta'x}}$$

os parâmetros estimados são apresentados no Quadro 11.

Quadro 11 — Modelo ajustado usando WOE

Variável	Categoria	$\hat{\beta}$	<i>p</i> -valor	$\hat{\beta} * WOE$	VIF
	<i>Intercepto</i>	0,19446	0,00000	0,19446	
Parcelamento	1) Sim	1,16938	0,00000	-6,52420	1,41318
	2) Não			1,23710	
Internet	2) Não	-0,25748	0,00004	-0,11417	2,31454
	1) Sim			0,15796	
FlagFraude	2) Não	0,24295	0,00001	0,02145	1,23235
	1) Sim			-0,47082	

FlagDesc	2) Não	2,69746	0,00000	-0,15850	1,17279
	1) Sim			1,93568	
StatusInst	Ativo parcial	-0,37980	0,00000	0,27631	1,50050
	Ativo total			-0,18403	
TipoTecnologia	ANTENA	-0,40797	0,00000	0,28895	1,51787
	CABO			-0,06281	
VIFatura	01. ≤ 20,62	0,89690	0,00000	0,64684	1,68712
	02. > 20,62 e ≤ 82,25			-0,73566	
	03. > 82,25 e ≤ 92,19			-0,42801	
	04. > 92,19 e ≤ 111,33			-0,12648	
	05. > 111,33 e ≤ 132,38			0,06784	
	06. > 132,38 e ≤ 171,59			0,26652	
	07. > 171,59			0,47335	
VIAjuste	01. ≤ 1,49	1,03176	0,00000	-0,00433	1,00824
	02. > 1,49 e ≤ 5,28			0,50297	
	03. > 5,28			-0,41668	
MediaPago	00. Missing	0,50078	0,00000	-1,41221	1,74263
	01. ≤ 15,00			-0,99343	
	02. > 15,00 e ≤ 31,55			-0,50139	
	03. > 31,55 e ≤ 83,19			-0,25235	
	04. > 83,19 e ≤ 90,71			0,00977	
	05. > 90,71 e ≤ 97,48			0,19742	
	06. > 97,48 e ≤ 110,30			0,35169	
	07. > 110,30 e ≤ 222,56			0,47322	
08. > 222,56	0,03696				
VIFatAtrasoMax	01. ≤ 145,08	0,88954	0,00000	0,40611	1,68331
	02. > 145,08 e ≤ 172,02			0,13570	
	03. > 172,02 e ≤ 214,84			-0,27569	
	04. > 214,84 e ≤ 250,43			-0,67407	
	05. > 250,43 e ≤ 313,86			-0,95479	
	06. > 313,86 e ≤ 455,93			-1,47175	
	07. > 455,93			-2,26959	
SeqFatura	01. ≤ 9	0,95820	0,00000	-1,37971	2,09784
	02. > 9 e ≤ 10			-0,15774	
	03. > 10 e ≤ 12			0,57430	
	04. > 12 e ≤ 14			1,14526	
	05. > 14 e ≤ 17			2,16110	
	06. > 17			3,72300	
L1feitos	01. ≤ 96,55	1,83274	0,00000	-0,66411	1,28405
	02. > 96,55 e ≤ 163,95			-0,12686	
	03. > 163,95 e ≤ 236,58			0,30256	

	04. > 236,58			0,72873	
L7feitos	01. ≤ 3,37	0,96165	0,00000	0,92815	1,74396
	02. > 3,37 e ≤ 6,69			0,74169	
	03. > 6,69 e ≤ 10,91			0,49244	
	04. > 10,91 e ≤ 16,12			0,22916	
	05. > 16,12 e ≤ 27,72			-0,03754	
	06. > 27,72 e ≤ 33,52			-0,24221	
	07. > 33,52 e ≤ 68,46			-0,57325	
	08. > 68,46 e ≤ 100,67			-0,98821	
	09. > 100,67 e ≤ 186,54			-1,57932	
	10. > 186,54			-2,55933	
L9feitos	01. ≤ 8,73	0,52532	0,00000	-0,15181	1,12672
	02. > 8,73 e ≤ 12,02			0,00330	
	03. > 12,02 e ≤ 29,96			0,11163	
	04. > 29,96 e ≤ 52,86			0,04500	
	05. > 52,86			-0,11145	
L10feitos	01. ≤ 2,78	1,11580	0,00000	0,14958	1,07434
	02. > 2,78 e ≤ 5,99			0,05500	
	03. > 5,99			-0,22174	
L21feitos	01. ≤ 0,32	0,48981	0,00010	-0,10782	1,14798
	02. > 0,32 e ≤ 0,78			-0,03960	
	03. > 0,78 e ≤ 40,18			0,07511	
	04. > 40,18			0,11865	
L20recebidos	01. Equal to zero	0,77565	0,00000	-0,02600	1,02766
	02. > 0,00 e ≤ 0,87			0,30292	
	03. > 0,87 e ≤ 2,16			0,06952	
	04. > 2,16 e ≤ 7,35			-0,27006	
	05. > 7,35 e ≤ 12,98			-0,04778	
	06. > 12,98			0,27847	
GROUPING - CL_CEP_C	01-Excelente/Muito Bom	0,59609	0,00000	0,31956	1,07522
	03-Bom			0,14614	
	04-Neutro			0,01278	
	05-Mau			-0,13757	
	06-Muito mau			-0,27662	
	07-Horrível/Missing			-0,48692	
GROUPING - CL_DsProdVoz	01-Excelente	0,42592	0,00000	0,44913	1,77659
	02-Muito Bom			0,24262	
	04-Neutro			-0,14331	
	05-Mau			-0,13973	
	06-Muito mau/Horrível/Missing			-0,51817	
GROUPING -	BCENT	1,02855	0,00000	-0,12627	1,02301

CanalVendas	Dealers			0,03888	
	Outros			0,45673	
GROUPING - Migracao	1) Sim	1,00540	0,00000	0,22848	1,12231
	2) Não			-0,21481	
GROUPING - WhiteList	00-Missing	0,69564	0,00000	-0,46528	1,19958
	01 - 0 a 2			1,29157	
	02 - 3 a 5			0,46491	
	03 - Acima de 5			-0,23742	

Fonte: a autora

Observa-se que o modelo ajustado apresenta ajuste satisfatório. Todos os parâmetros são significativos com valor-p $p < 0,05$, considerando nível de significância de 0,05. Após o ajuste do modelo, obteve-se, também, o quadro da Análise do Desvio para mensurar o desvio dos valores ajustados em relação aos dados observados.

Quadro 12 — Análise do Desvio - Modelo WOE

Fonte de variação	<i>Deviance</i>	Graus de liberdade	p-valor
Modelo	46.194,0	22	0,0000
Resíduos	13.698,9	43.316	1,0000
Total	59.892,9	43.338	

Fonte: a autora

Como o valor-p para o modelo ajustado na Análise do Desvio é $p = 0,0000 < 0,05$, pode-se concluir que existe uma relação estatisticamente significativa entre a variável resposta Y e as variáveis explicativas que compõem o vetor x. Além disso, o valor-p para os resíduos é $p = 1$, o que indica que o modelo não é significativamente pior do que o melhor modelo possível para esses dados. Com base nesses resultados, também é possível observar que a porcentagem do Desvio explicada pelo modelo é igual a 77,12%. Esta estatística é semelhante ao R^2 da regressão linear.

O modelo final ajustado usando o WOE como variável preditora aponta como principais fatores de risco de inadimplência basicamente os mesmos já identificados por meio do modelo com variáveis *dummies*: os parcelamentos, os alertas de fraude, os altos valores de fatura, os baixos valores de pagamentos históricos, os altos valores da fatura paga com o maior número de dias em atraso, o tempo de relacionamento (clientes novos representam maior risco), determinadas regiões identificadas por estado, cidade e CEP, além do tipo de serviço contratado e do perfil de consumo.

Para visualizar os resultados do modelo quanto a classificação dos clientes, o gráfico que acumula a quantidade de clientes Bons e Maus de acordo com a probabilidade é adequado. Observando o gráfico é possível visualizar que os clientes maus estão concentrados nas baixas probabilidades de pagamento, enquanto que os clientes bons se concentram nas altas probabilidades de pagamento.

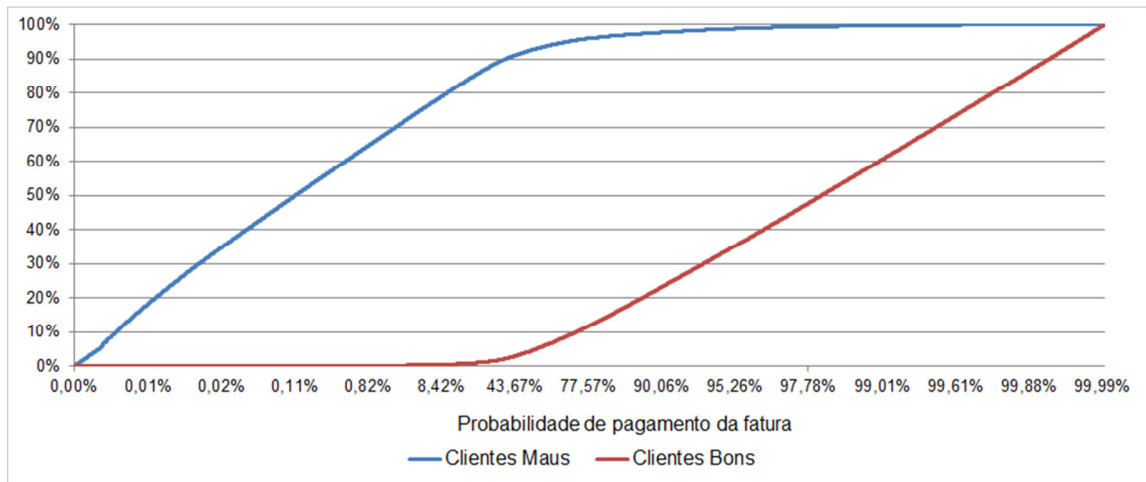


Figura 5 — Distribuição acumulada para os clientes Bons e Maus pela probabilidade de pagamento da fatura — Modelo WOE

4.6. COMPARAÇÃO DOS MODELOS NA AMOSTRA DE DESENVOLVIMENTO

A avaliação dos modelos foi feita construindo-se a matriz de confusão que compara o valor observado com o predito pelo modelo. Para o valor predito, foi considerado que, quando a probabilidade é maior que 0,5 de pagamento, o cliente é bom; caso contrário é classificado como mau.

Tabela 12 — Matriz de confusão para o modelo com variáveis *dummies* — com intercepto — Amostra de desenvolvimento

Observado	Predito		Total
	Mau	Bom	
Mau	18.827 93,0%	1.417 7,0%	20.244
Bom	725 3,1%	22.370 96,9%	23.095
Total	19.552 45,1%	23.787 54,9%	43.339

Fonte: a autora

Tabela 13 — Matriz de confusão para o modelo com variáveis *dummies* — sem intercepto — Amostra de desenvolvimento

Observado	Predito				Total
	Mau		Bom		
Mau	18.824	93,0%	1.420	7,0%	20.244
Bom	720	3,1%	22.375	96,9%	23.095
Total	19.544	45,1%	23.795	54,9%	43.339

Fonte: a autora

Tabela 14 — Matriz de confusão para o modelo usando WOE — Amostra de desenvolvimento

Observado	Predito				Total
	Mau		Bom		
Mau	18.619	92,0%	1.625	8,0%	20.244
Bom	887	3,8%	22.208	96,2%	23.095
Total	19.506	45,0%	23.833	55,0%	43.339

Fonte: a autora

Além da matriz de confusão, para comparação dos modelos, foram calculadas algumas estatísticas que mensuram o poder preditivo do modelo e são apresentadas na Tabela 15. Com base nesses indicadores, observa-se que o modelo ajustado utilizando as variáveis *dummies* como preditoras tem poder de discriminação levemente melhor, pois apresenta os índices de KS e Gini maior que o modelo ajustado utilizando o WOE como variável preditora. Outros indícios de que o modelo que utiliza variáveis *dummies* é levemente melhor é que ele apresenta menor taxa de erro de classificação, maiores valores preditivos, além de maiores índices de sensibilidade e especificidade.

Tabela 15 — Comparação dos resultados dos modelos na amostra de desenvolvimento

Estatística	Modelo <i>dummy</i> com intercepto	Modelo <i>dummy</i> sem intercepto	Modelo WOE
KS — índice de Kolmogorov-Smirnov	89,89%	89,90%	88,21%
Coeficiente de Gini	0,968	0,968	0,959
Taxa aparente de erro	4,94%	4,94%	5,80%
Valor preditivo do modelo	95,06%	95,06%	94,20%
Valor preditivo do modelo positivo	94,04%	94,03%	93,18%
Valor preditivo do modelo negativo	96,29%	96,32%	95,45%
Sensibilidade	96,86%	96,88%	96,16%
Especificidade	93,00%	92,99%	91,97%

Fonte: a autora

4.7. VALIDAÇÃO DO MODELO PELO MÉTODO DE LACHENBRUCH

Uma vez que a metodologia aplicada é de Reconhecimento de Padrões deve-se avaliar a sua eficiência no reconhecimento e classificação. Assim, aplicou-se o método de Lachenbruch descrito no item 2.4.9. O método foi programado no software SAS e aplicado ao grupo de clientes objeto deste estudo. Foram calculadas as matrizes de confusão, a taxa aparente de erro e a proporção classificada corretamente pelo modelo.

Tabela 16 — Matriz de confusão para o modelo usando variáveis *dummies* — Com intercepto — Obtido por Lachenbruch

Observado	Predito				Total
	Mau		Bom		
Mau	18.809	92,9%	1.435	7,1%	20.244
Bom	735	3,2%	22.360	96,8%	23.095
Total	19.544	45,1%	23.795	54,9%	43.339

Fonte: a autora

O modelo ajustado usando variáveis *dummies* com intercepto apresentou proporção de clientes classificada corretamente pelo modelo de 94,99%. Esta é a medida da eficiência do modelo ajustado.

Tabela 17 — Matriz de confusão para o modelo usando variáveis *dummies* — Sem intercepto — Obtido por Lachenbruch

Observado	Predito				Total
	Mau		Bom		
Mau	18.817	93,0%	1.427	7,0%	20.244
Bom	736	3,2%	22.359	96,8%	23.095
Total	19.544	45,1%	23.795	54,9%	43.339

Fonte: a autora

O modelo ajustado usando variáveis *dummies* sem intercepto apresentou proporção de clientes classificada corretamente pelo modelo de 95,01%. Esta é a medida da eficiência do modelo ajustado.

Tabela 18 — Matriz de confusão para o modelo usando WOE — Obtido por Lachenbruch

Observado	Predito				Total
	Mau		Bom		
Mau	18.617	92,0%	1.627	8,0%	20.244
Bom	872	3,8%	22.223	96,2%	23.095
Total	19.489	45,0%	23.850	55,0%	43.339

Fonte: a autora

O modelo ajustado usando o WOE apresentou proporção de clientes classificada corretamente pelo modelo de 94,23%. Esta é a medida da eficiência do modelo ajustado.

Os resultados observados no método de validação de Lachenbruch comprovam o que foi observado na amostra de desenvolvimento. De acordo com as estatísticas calculadas, o modelo ajustado utilizando variáveis *dummies* é levemente superior ao modelo ajustado com o WOE.

5. CONSIDERAÇÕES FINAIS E CONCLUSÃO

Neste trabalho, analisou-se um conjunto de dados visando a identificação dos principais fatores de risco para a inadimplência. Ajustou-se modelos preditivos para a classificação de clientes com base na probabilidade de pagamento das faturas mensais de uma empresa da área de telecomunicações, comparou-se esses modelos visando identificar o que tem melhor desempenho. Todos os passos para a obtenção de um bom modelo de previsão foram apresentados, desde a definição do conjunto de dados, tratamento das variáveis e ajuste do modelo final.

Para segmentar a base de dados, foi aplicada a Análise de *Cluster*. Com base nos grupos formados por meio dos métodos hierárquicos e não hierárquicos, conclui-se que a amostra é homogênea, ou seja, há alta concentração de clientes num único *cluster*. Portanto, a amostra foi tratada como um grupo único.

Para o tratamento, as variáveis foram divididas em dois grupos, conforme suas características, sendo um grupo de variáveis categóricas e outro grupo de variáveis contínuas. As classes para essas variáveis foram criadas baseando-se no risco relativo e seu intervalo de confiança.

Para ajuste do modelo, foi utilizada a regressão logística. As variáveis preditoras foram incluídas no modelo de duas formas, a primeira considerando variáveis *dummies* e a segunda considerando o peso de evidência (WOE — *Weight of Evidence*).

Nas duas formas de análise, com as variáveis *dummies* ou usando o WOE, foram identificados como principais fatores de risco de inadimplência: os parcelamentos, os alertas de fraude, os altos valores de fatura, os baixos valores de pagamentos históricos, os altos valores da fatura paga com o maior número de dias em atraso, o tempo de relacionamento (clientes novos representam maior risco), determinadas regiões identificadas por meio de estado, cidade e CEP, além do tipo de serviço contratado e do perfil de consumo.

Ao comparar o desempenho dos modelos na amostra de desenvolvimento, observou-se que o poder preditivo do modelo ajustado utilizando as variáveis *dummies* como preditoras é levemente melhor, pois apresenta os índices de KS e Gini maior que o modelo ajustado utilizando o WOE como variável preditora. Além disso, apresenta

menor taxa de erro de classificação, maiores valores preditivos e maiores índices de sensibilidade e especificidade.

Foi aplicado o método de Lachenbruch para validação e para avaliar a eficiência dos modelos. Na validação, os resultados confirmaram a superioridade do modelo desenvolvido utilizando as variáveis *dummies*, embora pequena. Outro ponto importante é que, em ambos os modelos, não houve redução significativa do poder de discriminação do modelo ao se comparar os resultados observados na amostra de desenvolvimento com a de validação. Isso é importante, pois indica que os modelos não perdem poder de discriminação ao serem aplicados num registro que não estava presente no desenvolvimento do modelo. Portanto os modelos podem ser utilizados para novos clientes que contratarem os produtos da empresa. Esses podem facilmente ser implantados num sistema automático para cálculo do *score*.

Do ponto de vista de desenvolvimento de modelos, é mais simples trabalhar com o que considera o WOE como variável preditora, pois isso reduz a dimensão do problema, uma vez que não será necessário criar $k-1$ variáveis *dummies* para as classes de cada variável. Além disso, esse modelo reduz a quantidade de parâmetros a ser estimada, pois se estima um parâmetro para cada variável, e não uma para cada *dummy*. Isso é particularmente interessante ao se trabalhar com bases de dados que contêm muitas variáveis.

É importante mencionar que não se pode afirmar que um cliente classificado pelo modelo como bom pagador certamente pagará a fatura, mas sua probabilidade de pagamento é maior que de um cliente classificado como mau.

Como sugestão para trabalhos futuros relacionados com esse tema, propõe-se que sejam avaliadas outras estatísticas além do WOE para utilizar, substituindo as variáveis *dummies*, além da aplicação de outros métodos estatísticos multivariados, como árvores de decisão ou redes neurais.

REFERÊNCIAS

ANDRADE, F. W. (2004). **Desenvolvimento de modelo de Risco de Portfólio para Carteiras de Crédito a Pessoas Físicas**. *Tese de Doutorado*. São Paulo, SP, Brasil.

BARROSO, L. P., & ARTES, R. (2003). **Análise Multivariada - Minicurso 2**. Lavras: UFLA.

CHAVES NETO, A. (2007). **Análise Multivariada II**. *Notas de aula da disciplina Análise Multivariada II*. Curitiba, Paraná, Brasil.

GIOLO, S. R. (2006). **Introdução a Análise de dados Categóricos**. *Notas de aula da disciplina Introdução a Análise de dados Categóricos*. Curitiba, Paraná, Brasil.

GIOLO, S. R. (2011). **Análise de Regressão Linear**. *Análise de Regressão Linear*. Curitiba, PR, Brasil.

GONÇALVES, E. B. (2005). **Análise do Risco de Crédito com o uso de modelos de Regressão Logística, Redes Neurais e Algoritmos Genéticos**. *Dissertação de Mestrado*. São Paulo, São Paulo, Brasil.

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

IDEC - Instituto Brasileiro de Defesa do Consumidor. (s.d.). *IDEC*. Acesso em 24 de Outubro de 2011, disponível em IDEC: <http://www.idec.org.br/>

JOHNSON, R. A., & WICHERN, D. W. (1998). **Applied Multivariate Statistical Analysis**. Simon & Schuster.

MANFIO, F. (2007). **O Risco Nosso de Cada Dia**. Estação das Letras e Cores.

OHTOSHI, C. (2003) **Uma Comparação de Regressão Logística, Árvores de Classificação e Redes Neurais: Analisando Dados de Crédito.** Dissertação de Mestrado. Departamento de Estatística. Universidade de São Paulo. IME/USP.

PAULA, G. A. (2010). **Modelos de Regressão com apoio computacional.** <http://www.ime.usp.br/~giapaula/mlgs.html>. São Paulo, SP, Brasil.

RELIASOFT. (2011). *ReliaSoft*. Acesso em 24 de Outubro de 2011, disponível em Relia Soft Brasil - Conceitos de Confiabilidade: Estimador de Máxima Verossimilhança: <http://www.reliasoft.com.br/hotwire/edicao2/conceito2.htm>

ROSA, P. (2000). **Modelos de Credit Scoring: Regressão Logística, CHAID e REAL.** *Dissertação de Mestrado.* São Paulo, SP, Brasil.

SHARMA, S. (1996). ***Applied Multivariate Techniques.*** John Wiley & Sons, Inc.

SICSÚ, A. L. (2010). ***Credit Scoring: desenvolvimento, implantação, acompanhamento.*** São Paulo: Edgard Blucher.

TELECO - Inteligência em Telecomunicações. (s.d.). *Teleco*. Acesso em 24 de Outubro de 2011, disponível em Teleco: <http://www.teleco.com.br/>

THOMAS, L. C. (2000). **A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers.** *International Journal of Forecasting*.

APÊNDICES

Apêndice 1 – Fusão das classes das variáveis categóricas

CEP

<i>Categoria</i>	<i>% Total</i>	<i>RR</i>	<i>Bom</i>	<i>Mau</i>	<i>% Bons</i>	<i>IC 95% Inf.</i>	<i>IC 95% Sup.</i>	<i>Classe</i>
00-Missing	0,02	0,15	1	6	0,14	-0,21	0,49	00-Missing
03-REG.LESTE - SP	0,01	3,51	4	1	0,8	0,24	1,36	04-Neutro
04-REG.SUL - SP	0,04	4,67	16	3	0,84	0,66	1,02	04-Neutro
06-REG.OSASCO - SP	0,01	0,88	2	2	0,5	-0,42	1,42	04-Neutro
09-REG. ABCD - SP	0,01	0,88	2	2	0,5	-0,42	1,42	04-Neutro
12-VALE DO PARAÍBA	0,01	1,75	4	2	0,67	0,12	1,21	04-Neutro
13-REG.CAMPINAS	0,03	8,77	10	1	0,91	0,71	1,11	04-Neutro
14-REG.RIBEIRO PRETO	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
15-REG.S.J.RIO PRETO	0,01	0,29	1	3	0,25	-0,55	1,05	04-Neutro
18-REG. SOROCABA	0,02	2,63	6	2	0,75	0,36	1,14	04-Neutro
19-REG. PRES.PRUDENTE	0,01	0,88	2	2	0,5	-0,42	1,42	04-Neutro
20-CENTRO - RJ	0,01	0,88	3	3	0,5	-0,07	1,07	04-Neutro
22-RJ - SUL	0,03	3,94	9	2	0,82	0,55	1,09	04-Neutro
29-ESPIRITO SANTO	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro
30-BELO HORIZONTE	1,64	2,39	519	190	0,73	0,7	0,76	01-Excelente
31-BELO HORIZONTE	0,48	1,22	120	86	0,58	0,51	0,65	04-Neutro
38-REG. UBERABA - MG	0,02	2,19	5	2	0,71	0,26	1,17	04-Neutro
40-SALVADOR - BA	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
41-SALVADOR - BA	0,01	1,75	2	1	0,67	-0,77	2,1	04-Neutro
66-BELEM - PA	0,01	2,63	3	1	0,75	-0,05	1,55	04-Neutro
69-MANAUS - AM	0,09	0,13	5	33	0,13	0,02	0,24	04-Neutro
70-BRASILIA - DF	3,11	1,58	869	481	0,64	0,62	0,67	02-Muito bom
71-BRASILIA - DF	2,36	0,98	540	483	0,53	0,5	0,56	04-Neutro
72-TAGUATINGA - DF	5,8	0,66	1.083	1.429	0,43	0,41	0,45	06-Muito Mau
73-SOBRADINHO - DF	0,29	1,01	68	59	0,54	0,45	0,62	04-Neutro
74-GOIANIA - GO	7,5	0,71	1.450	1.800	0,45	0,43	0,46	05-Mau
75-ANAPOLIS - GO	1,04	0,51	166	284	0,37	0,32	0,41	06-Muito Mau
76-S. LUIS MTES CLAROS - GO	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
77-CAMPOS BELOS - GO	0,13	0,18	10	48	0,17	0,07	0,27	04-Neutro
78-CUIABA - MT	3,83	0,48	590	1.069	0,36	0,33	0,38	07-Horrível

79-CAMPO GRANDE - MS	4,43	0,89	966	955	0,5	0,48	0,53	05-Mau
80-CURITIBA - PR	6,6	1,55	1.827	1.035	0,64	0,62	0,66	02-Muito bom
81-CURITIBA - PR	5,81	1,01	1.346	1.170	0,53	0,52	0,55	04-Neutro
82-CURITIBA - PR	3,79	1,06	901	743	0,55	0,52	0,57	04-Neutro
83-REND.CURITIBA - PR	2,04	0,67	384	501	0,43	0,4	0,47	06-Muito Mau
84-PONTA GROSSA - PR	1,67	0,98	382	342	0,53	0,49	0,56	04-Neutro
85-GUARAPUAVA - PR	4,37	1,15	1.076	817	0,57	0,55	0,59	03-Bom
86-LONDRINA - PR	4,48	0,9	981	959	0,51	0,48	0,53	05-Mau
87-MARINGA - PR	6,85	1,3	1.772	1.195	0,6	0,58	0,61	03-Bom
88-FLORIANOPOLIS - SC	4,9	1,48	1.336	789	0,63	0,61	0,65	03-Bom
89-BLUMENAU - SC	3,54	1,45	958	578	0,62	0,6	0,65	03-Bom
90-PORTO ALEGRE - RS	6,2	1,14	1.518	1.171	0,56	0,55	0,58	03-Bom
91-PORTO ALEGRE - RS	3,93	0,59	686	1.019	0,4	0,38	0,43	06-Muito Mau
92-CANOAS - RS	2,88	0,74	573	675	0,46	0,43	0,49	05-Mau
93-SAO LEOPOLDO - RS	2,31	0,8	478	522	0,48	0,45	0,51	05-Mau
94-GRAVATAI - RS	2,14	0,38	278	649	0,3	0,27	0,33	07-Horriavel
95-CAXIAS DO SUL	2,75	1,93	818	372	0,69	0,66	0,71	02-Muito bom
96-PELOTAS - RS	1,92	1,62	540	292	0,65	0,62	0,68	02-Muito bom
97-SANTA MARIA - RS	2,01	1,21	505	365	0,58	0,55	0,61	03-Bom
99-PASSO FUNDO - RS	0,72	2,19	222	89	0,71	0,66	0,76	01-Excelente

Natureza de Ocupação

<i>Categoria</i>	<i>% Total</i>	<i>RR</i>	<i>Bom</i>	<i>Mau</i>	<i>% Bons</i>	<i>IC 95% Inf.</i>	<i>IC 95% Sup.</i>	<i>Classe</i>
0-Espólio	40,62	0,89	8.866	8.737	0,5	0,5	0,51	05-Mau
00-Missing	1,76	1,61	495	269	0,65	0,61	0,68	00-Missing
1-Empregado Empresa Setor Privado	7,02	1,08	1.681	1.363	0,55	0,53	0,57	04-Neutro
2-Profissional Liberal ou Trab s/ vínculo	17,96	1,01	4.173	3.609	0,54	0,53	0,55	04-Neutro
3-Empregador-titular ou Proprietário Empresa	8,66	0,91	1.913	1.839	0,51	0,49	0,53	04-Neutro
4-Serviço Público da Adm Direta	12,99	1,17	3.217	2.412	0,57	0,56	0,58	03-Bom
5-Serviço Público de Autarquia e Fund	5,81	0,92	1.290	1.227	0,51	0,49	0,53	04-Neutro
6-Funcionário Empresa Pública ou Economia Mista	0,15	2,12	46	19	0,71	0,59	0,82	04-Neutro
7-Rendimento Capital/Aluguéis	0,51	1,55	140	79	0,64	0,58	0,7	02-Muito bom

8-Aposentado/Pensionista	0,96	1,36	253	163	0,61	0,56	0,66	03-Bom
9-Outros	3,57	1,7	1.021	527	0,66	0,64	0,68	02-Muito bom

OCUPACAO

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
00-Missing	45,85	0,88	9.966	9.904	0,5	0,49	0,51	00-Missing
101-ENGENHEIRO	0,99	3,47	344	87	0,8	0,76	0,84	01-Excelente
102-ARQUITETO	0,36	1,67	103	54	0,66	0,58	0,73	04-Neutro
103-AGRONOMO	0,27	1,01	62	54	0,53	0,44	0,63	04-Neutro
104-QUIMICO	0,05	0,8	11	12	0,48	0,26	0,7	04-Neutro
105-DESENHISTA INDUSTRIAL	0,03	0,73	5	6	0,45	0,1	0,81	04-Neutro
106-FISICO	0,01	0,29	1	3	0,25	-0,55	1,05	04-Neutro
107-GEOLOGO	0,07	1,24	17	12	0,59	0,4	0,78	04-Neutro
108-TECNOLOGO	0,06	4,82	22	4	0,85	0,7	0,99	04-Neutro
111-MEDICO	1,32	2	398	174	0,7	0,66	0,73	01-Excelente
113-ENFERMEIRO E NUTRICIONISTA	0,46	0,74	92	109	0,46	0,39	0,53	04-Neutro
114-FISIOTERAPEUTA /TERAPEUTA OCUPAC	0,13	2,11	41	17	0,71	0,59	0,83	04-Neutro
115-ODONTOLOGO	0,8	2,33	252	95	0,73	0,68	0,77	01-Excelente
116-BIOLOGO E BIOMEDICO	0,18	0,88	39	39	0,5	0,39	0,61	04-Neutro
117-FARMACEUTICO	0,42	1,28	108	74	0,59	0,52	0,67	04-Neutro
121-ECONOMISTA	0,23	1,41	61	38	0,62	0,52	0,71	04-Neutro
123-ATUARIO E MATEMATICO	0,01	0	0	6	0	0	0	04-Neutro
124-CONTADOR	0,49	2,27	153	59	0,72	0,66	0,78	04-Neutro
125-ADMINISTRADOR	13,3	0,99	3.059	2.704	0,53	0,52	0,54	04-Neutro
127-GEOGRAFO	0,02	1,31	6	4	0,6	0,23	0,97	04-Neutro
128-ASTRONOMO E METEOROLOGISTA	0,01	1,31	3	2	0,6	-0,08	1,28	04-Neutro
129-OUTROS ECONOM ESTAT CONT	0,16	0,72	31	38	0,45	0,33	0,57	04-Neutro
131-ADVOGADO	4,57	1,41	1.222	757	0,62	0,6	0,64	03-Bom
132-PSICOLOGO	0,17	3,82	61	14	0,81	0,72	0,9	04-Neutro
133-SOCIOLOGO	0,03	0,88	6	6	0,5	0,17	0,83	04-Neutro
134-ASSISTENTE SOCIAL	0,07	2,52	23	8	0,74	0,58	0,91	04-Neutro
135-BIBLIOT/ARQUIV/ MUSEOL/ARQUEOL	0,52	0,57	89	138	0,39	0,33	0,46	06-Muito Mau
136-COMUNICOLOGO	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro

138-PROFISSIONAIS DE LETRAS E ARTES	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro
142-PROFESSOR DE ENSINO SUP.	0,69	2,74	228	73	0,76	0,71	0,81	01-Excelente
143-PROFESSOR ENS PRIM E SEG GRAUS	0,78	1,54	215	122	0,64	0,59	0,69	02-Muito bom
144-DIRETOR ESTABELEC ENSINO	0,24	1,4	64	40	0,62	0,52	0,71	04-Neutro
145-OUTROS PROF E ASSEMELHADOS	0,01	1,75	2	1	0,67	-0,77	2,1	04-Neutro
151-TECNICO CONTAB E ESTATISTICA	0,08	1,05	18	15	0,55	0,37	0,72	04-Neutro
152-TECNICO DE BIOLOGIA	0,02	0,15	1	6	0,14	-0,21	0,49	04-Neutro
153-TECNICO AGRONOMIA E AGRIMENSURA	0,01	1,31	3	2	0,6	-0,08	1,28	04-Neutro
154-TECNICO DE QUIMICA	0,03	1,75	8	4	0,67	0,35	0,98	04-Neutro
155-TECNICO DE MECANICA	0,1	1,17	24	18	0,57	0,42	0,73	04-Neutro
156-TECNICO ELETRIC/ELETRON/TELECOM	0,31	1,03	72	61	0,54	0,46	0,63	04-Neutro
157-TECNICO DE LABORATORIO E RAIOS X	0,11	0,99	26	23	0,53	0,39	0,68	04-Neutro
158-DESENHISTA TECNICO	0,05	0,8	11	12	0,48	0,26	0,7	04-Neutro
161-EMPRESARIO E PRODUTOR	0,21	0,82	45	48	0,48	0,38	0,59	04-Neutro
163-CANTOR E COMPOSITOR	0,01	0,18	1	5	0,17	-0,26	0,6	04-Neutro
164-MUSICO	0,1	0,69	19	24	0,44	0,29	0,6	04-Neutro
165-COREOGRAFO E BAILARINO	0,01	1,75	4	2	0,67	0,12	1,21	04-Neutro
166-LOCUTOR/COMENTARISTA/RADIALISTA	0,07	0,46	10	19	0,34	0,16	0,53	04-Neutro
168-ATLETA PROFISS E TECNICO DESPORT	0,04	1,25	10	7	0,59	0,33	0,85	04-Neutro
169-OUTROS TRAB ESPETAC PUBLICOS	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
171-JORNALISTA	0,12	1,75	34	17	0,67	0,53	0,8	04-Neutro
172-PUBLICITARIO	0,2	1,03	46	39	0,54	0,43	0,65	04-Neutro
182-COMISSARIO DE BORDO	0,01	2,63	3	1	0,75	-0,05	1,55	04-Neutro
189-OUTROS TRAB LIGADOS A NAVEGACAO	0,08	1,04	19	16	0,54	0,37	0,72	04-Neutro
191-ESCUPTOR/PINTOR E ASSEMELHADOS	0,05	1,36	14	9	0,61	0,39	0,82	04-Neutro
192-DESENHISTA COMERCIAL	0,03	1,05	6	5	0,55	0,19	0,9	04-Neutro
193-DECORADOR	0,03	0,73	5	6	0,45	0,1	0,81	04-Neutro
199-OUTROS TRAB PROFISS CIENTIF	0,13	0,57	22	34	0,39	0,26	0,52	04-Neutro
201-LEG: SENAD/DEP FED/DEP EST/VEREA	0,02	1,46	5	3	0,63	0,19	1,06	04-Neutro
205-JUDIC: MIN TRIB	0,08	0,82	16	17	0,48	0,3	0,66	04-Neutro

SUP/DESEMB/JUIZ									
211-PROCURADOR DE ASSEMBLHADOS	0,1	1,44	28	17	0,62	0,47	0,77	04-Neutro	
212-DIPLOMATA	0,03	0,5	4	7	0,36	0,02	0,7	04-Neutro	
216-OFFICIAL FORCAS ARM/FORCAS AUXIL	0,09	4,82	33	6	0,85	0,73	0,96	04-Neutro	
293-SERVENTUARIO DE JUSTICA	0,08	2,74	25	8	0,76	0,6	0,91	04-Neutro	
294-TABELIAO	0,21	0,79	43	48	0,47	0,37	0,58	04-Neutro	
295-MILITAR EM GERAL	0,12	0,84	25	26	0,49	0,35	0,63	04-Neutro	
296-SERVIDOR PUBLICO FEDERAL	3,19	1,43	856	526	0,62	0,59	0,65	03-Bom	
297-SERVIDOR PUBLICO ESTADUAL	0,74	1,64	208	111	0,65	0,6	0,7	02-Muito bom	
298-SERVIDOR PUBLICO MUNICIPAL	0,18	1,4	48	30	0,62	0,5	0,73	04-Neutro	
299-OUTROS SERV CIVIS E MILITARES	1,14	0,98	261	234	0,53	0,48	0,57	04-Neutro	
301-DIRETOR DE EMPRESAS	0,09	0,88	20	20	0,5	0,34	0,66	04-Neutro	
309-OUTROS DIRETORES E GERENTES	0,1	1,17	24	18	0,57	0,42	0,73	04-Neutro	
391-CHEFE INTERMEDIARIO	0,7	0,97	159	144	0,52	0,47	0,58	04-Neutro	
392-TRABAL SERV CONTAB/CAIXA/ASSEMEL	0,09	1,07	22	18	0,55	0,39	0,71	04-Neutro	
393-SECR/ESTEN/DATIL/RECEP/TE LEF/ASS	0,61	0,5	96	168	0,36	0,31	0,42	06-Muito Mau	
394-AUXILIAR DE ESCRITORIO E ASSEMEL	0,11	0,51	18	31	0,37	0,23	0,51	04-Neutro	
395-BANCARIO E ECONOMIARIO	1,09	2,9	364	110	0,77	0,73	0,81	01-Excelente	
396-SECURITARIO	0,03	1,31	9	6	0,6	0,32	0,88	04-Neutro	
399-OUTR TRAB ADM E ASSEMBLHADOS	0,05	1,17	12	9	0,57	0,34	0,8	04-Neutro	
401-SUPERV/INSPET/AG COMPRA E VENDA	0,03	1,05	6	5	0,55	0,19	0,9	04-Neutro	
402-VEND PRAC/REPR COM/CAIX VIAJ/ASS	0,51	0,77	103	117	0,47	0,4	0,53	04-Neutro	
403-CORRET IMOV/SEGURO/TIT E VALORES	0,24	0,88	53	53	0,5	0,4	0,6	04-Neutro	
405-AGENCIADOR DE PROPAGANDA	0,06	0,55	10	16	0,38	0,18	0,59	04-Neutro	
411-VENDEDOR COMERCIO VAREJ/ATACAD	1,37	0,55	227	365	0,38	0,34	0,42	06-Muito Mau	
413-FEIRANTE	0,52	0,35	64	162	0,28	0,22	0,34	07-Horrivel	
419-OUTROS VENDEDORES E ASSEMBLHADOS	0,69	0,64	126	172	0,42	0,37	0,48	06-Muito Mau	
491-DEMONSTRADOR	0,01	0	0	5	0	0	0	04-Neutro	
499-OUTROS TRABALHADORES COMERCIO	0,06	0,36	7	17	0,29	0,1	0,49	04-Neutro	

501-PORTEIRO/ASCENS/GARAG/FAXINEIRO	0,07	0,46	10	19	0,34	0,16	0,53	04-Neutro
511-CABELEIR/BARB/MANIC/PEDIC/MAQUIL	0,5	0,42	70	146	0,32	0,26	0,39	04-Neutro
521-GOVERN HOTEL/CAMAR/PORT/COZ/GAR C	0,12	0,26	12	40	0,23	0,11	0,35	04-Neutro
529-OUTROS TRAB SERV HOSPEDAGEM	0,06	0,29	7	21	0,25	0,08	0,42	04-Neutro
531-MOTORISTA VEIC TRANSP PASSAG	0,16	0,39	22	49	0,31	0,2	0,42	04-Neutro
532-MOTORISTA VEIC TRANSP CARGA	0,1	0,34	12	31	0,28	0,14	0,42	04-Neutro
541-MECANICO SERV AUTOM/MAQUINAS	0,14	0,48	21	38	0,36	0,23	0,48	04-Neutro
542-ELETRIC SERV AUTOM/MAQ/APAR ELET	0,11	0,74	21	25	0,46	0,31	0,61	04-Neutro
543-LANTERN/PINTOR VEIC METALICOS	0,03	0,39	4	9	0,31	0,02	0,6	04-Neutro
591-ALFAIATE	0,02	2,19	5	2	0,71	0,26	1,17	04-Neutro
592-PROTETICO	0,03	1,17	8	6	0,57	0,27	0,87	04-Neutro
593-DESPACHANTE INCLUSIVE ADUANEIRO	0,04	0,88	9	9	0,5	0,24	0,76	04-Neutro
594-AGENTE VIAGEM E GUIA TURISMO	0,18	0,81	38	41	0,48	0,37	0,59	04-Neutro
596-AUXILIAR DE LABORATORIO	0,02	0,53	3	5	0,38	-0,06	0,81	04-Neutro
601-TRABALHADOR AGRICOLA	0,07	0,77	14	16	0,47	0,28	0,66	04-Neutro
602-TRABALHADOR DA PECUARIA	0,05	1,63	13	7	0,65	0,42	0,88	04-Neutro
701-MESTRE E CONTRAMESTRE	0,02	0,11	1	8	0,11	-0,15	0,37	04-Neutro
717-JOALHEIROS E OURIVES	0,03	0,44	5	10	0,33	0,06	0,6	04-Neutro
719-OUTROS ALUGUEIS VEIC MAQ	0,06	0,7	12	15	0,44	0,24	0,64	04-Neutro
901-PROPR ESTAB AGRIC PEC FLOREST	0,03	0,77	7	8	0,47	0,18	0,75	04-Neutro
902-PROPR ESTAB COMERCIAL	3,95	0,78	809	904	0,47	0,45	0,5	05-Mau
903-PROPR ESTAB INDUSTRIAL	0,05	3,94	18	4	0,82	0,64	0,99	04-Neutro
905-PROPR MICROEMPRESA	0,93	0,79	190	212	0,47	0,42	0,52	05-Mau
906-PROPR IMOVEL RECEB RENDIM ALUGUE	0,37	0,84	78	81	0,49	0,41	0,57	04-Neutro
907-CAPITALISTA RECEB REND APLIC CAP	0,01	0,88	3	3	0,5	-0,07	1,07	04-Neutro
921-MILITAR REFORMADO	0,22	4,27	78	16	0,83	0,75	0,91	04-Neutro
922-FUNCIONARIO PUBL CIVIL APOSENT	0,17	1,44	46	28	0,62	0,51	0,73	04-Neutro

923-APOSENTADO (EXC FUNC PUBL)	2,11	1,97	634	282	0,69	0,66	0,72	02-Muito bom
924-PENSIONISTA	0,37	1,03	86	73	0,54	0,46	0,62	04-Neutro
931-BOLSISTA ESTAGIARIO ASSEMBLHADO	0,99	1,42	266	164	0,62	0,57	0,66	03-Bom
999-OUTR DECLAR NAO ESPECIF ANTERIOR	2,5	1,16	618	467	0,57	0,54	0,6	03-Bom

TpResidencia

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
00-Missing	40,57	0,89	8.864	8.720	0,5	0,5	0,51	00-Missing
01-Propria	53,55	1,11	12.958	10.252	0,56	0,55	0,56	04-Neutro
02-Alugada	5,55	0,88	1.204	1.202	0,5	0,48	0,52	05-Mau
03-Cedida	0,32	0,86	69	70	0,5	0,41	0,58	04-Neutro

EstCivil

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
00-Missing	36,44	0,9	8.022	7.772	0,51	0,5	0,52	00-Missing
01-Casado	48,57	1,11	11.744	9.307	0,56	0,55	0,56	04-Neutro
02-Solteiro	11,63	0,86	2.499	2.542	0,5	0,48	0,51	05-Mau
03-Divorciado/Desquitado	1,89	1,12	458	360	0,56	0,53	0,59	03-Bom
04-Viúvo	1,14	1,56	317	178	0,64	0,6	0,68	02-Muito bom
05-Uniao Estavel	0,32	0,57	55	85	0,39	0,31	0,47	04-Neutro

UF

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
AC-Acre	0,09	0,13	5	33	0,13	0,02	0,24	04-Neutro
DF-Distrito Federal	11,62	0,92	2.575	2.460	0,51	0,5	0,53	04-Neutro
GO-Goiás	8,57	0,68	1.623	2.089	0,44	0,42	0,45	05-Mau
MG-Minas Gerais	2,11	2	635	279	0,69	0,66	0,72	02-Muito bom
MS-Mato Grosso do Sul	4,47	0,89	976	960	0,5	0,48	0,53	05-Mau
MT-Mato Grosso	3,48	0,53	566	942	0,38	0,35	0,4	06-Muito Mau
PR-Paraná	35,76	1,13	8.717	6.780	0,56	0,55	0,57	03-Bom
RO-Rondônia	0,36	0,18	26	130	0,17	0,11	0,23	04-Neutro
RS-Rio Grande do Sul	24,97	0,96	5.654	5.167	0,52	0,51	0,53	04-Neutro

SC-Santa Catarina	8,45	1,49	2.302	1.358	0,63	0,61	0,64	03-Bom
TO-Tocantins	0,13	0,17	9	46	0,16	0,06	0,26	04-Neutro

Cidade

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
ALMIRANTE TAMANDARE	0,1	0,4	14	31	0,31	0,17	0,45	04-Neutro
ALVORADA	0,31	0,23	28	105	0,21	0,14	0,28	04-Neutro
ANAPOLIS	1,03	0,51	165	282	0,37	0,32	0,41	06-Muito Mau
APARECIDA DE GOIANIA	0,79	0,41	110	233	0,32	0,27	0,37	07-Horrivel
APUCARANA	0,68	1,01	157	136	0,54	0,48	0,59	04-Neutro
ARAPONGAS	0,5	1,11	120	95	0,56	0,49	0,63	04-Neutro
ARAUCARIA	0,13	0,66	25	33	0,43	0,3	0,56	04-Neutro
BALNEARIO CAMBORIU	0,26	1,39	68	43	0,61	0,52	0,7	04-Neutro
BELO HORIZONTE	2,11	2	635	279	0,69	0,66	0,72	02-Muito bom
BENTO GONCALVES	0,21	3,71	72	17	0,81	0,73	0,89	04-Neutro
BLUMENAU	0,83	1,19	206	152	0,58	0,52	0,63	04-Neutro
BRASILIA	2,5	1,56	692	390	0,64	0,61	0,67	02-Muito bom
BRUSQUE	0,09	1,62	24	13	0,65	0,49	0,81	04-Neutro
CACHOEIRINHA	0,32	0,62	57	81	0,41	0,33	0,5	04-Neutro
CAMBE	0,18	0,88	38	38	0,5	0,38	0,62	04-Neutro
CAMPO BOM	0,07	0,62	12	17	0,41	0,22	0,6	04-Neutro
CAMPO GRANDE	4,47	0,89	976	960	0,5	0,48	0,53	05-Mau
CAMPO LARGO	0,2	1,11	48	38	0,56	0,45	0,67	04-Neutro
CAMPO MOURAO	0,52	1,37	138	88	0,61	0,55	0,67	03-Bom
CANDANGOLANDIA	0,02	0	0	7	0	0	0	04-Neutro
CANOAS	2,82	0,75	563	659	0,46	0,43	0,49	05-Mau
CASCAVEL	1,96	1,59	548	303	0,64	0,61	0,68	02-Muito bom
CAXIAS DO SUL	2,25	1,68	641	334	0,66	0,63	0,69	02-Muito bom
CEILANDIA	0,8	0,44	117	231	0,34	0,29	0,39	07-Horrivel
CIANORTE	0,48	1,47	129	77	0,63	0,56	0,69	04-Neutro
COLOMBO	0,41	0,52	66	111	0,37	0,3	0,44	04-Neutro
CRICIUMA	0,39	1,22	99	71	0,58	0,51	0,66	04-Neutro
CRUZEIRO	0,74	1,67	211	111	0,66	0,6	0,71	02-Muito bom
CUIABA	3,02	0,59	524	783	0,4	0,37	0,43	06-Muito Mau
CURITIBA	16,35	1,21	4.116	2.972	0,58	0,57	0,59	03-Bom
ESTEIO	0,22	0,47	33	62	0,35	0,25	0,44	04-Neutro
FARROUPILHA	0,06	3,33	19	5	0,79	0,62	0,97	04-Neutro

FLORIANOPOLIS	1,45	2,12	446	184	0,71	0,67	0,74	01-Excelente
FOZ DO IGUACU	1,89	0,77	383	434	0,47	0,43	0,5	05-Mau
GAMA	0,21	0,84	45	47	0,49	0,39	0,59	04-Neutro
GOIANIA	6,74	0,75	1.347	1.572	0,46	0,44	0,48	05-Mau
GRAVATAI	1,04	0,43	149	301	0,33	0,29	0,37	07-Horrivel
GUAIBA	0,09	0,83	18	19	0,49	0,32	0,66	04-Neutro
GUARA	1,27	1,01	295	255	0,54	0,49	0,58	04-Neutro
GUARAPUAVA	0,25	1	58	51	0,53	0,44	0,63	04-Neutro
ICARA	0,03	0,88	7	7	0,5	0,2	0,8	04-Neutro
ITAJAI	0,38	0,64	70	96	0,42	0,35	0,5	04-Neutro
JARAGUA DO SUL	0,15	4,02	55	12	0,82	0,73	0,92	04-Neutro
JOINVILLE	2,6	1,51	711	414	0,63	0,6	0,66	03-Bom
LONDRINA	2,55	0,75	511	595	0,46	0,43	0,49	05-Mau
LUZIANIA	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro
MARIALVA	0,3	1,43	80	49	0,62	0,54	0,71	04-Neutro
MARINGA	5,15	1,37	1.361	873	0,61	0,59	0,63	03-Bom
MONTENEGRO	0,23	5,38	86	14	0,86	0,79	0,93	04-Neutro
NOVO HAMBURGO	1,05	0,9	231	226	0,51	0,46	0,55	05-Mau
NUCLEO BANDEIRANTE	0,23	0,51	37	64	0,37	0,27	0,46	04-Neutro
PALHOCA	0,22	0,79	45	50	0,47	0,37	0,58	04-Neutro
PALMAS	0,13	0,17	9	46	0,16	0,06	0,26	04-Neutro
PARANAGUA	0,16	0,42	22	46	0,32	0,21	0,44	04-Neutro
PASSO FUNDO	0,71	2,19	220	88	0,71	0,66	0,77	01-Excelente
PELOTAS	1,37	1,51	375	217	0,63	0,59	0,67	02-Muito bom
PINHAIS	0,54	0,85	115	119	0,49	0,43	0,56	04-Neutro
PIRAQUARA	0,1	0,36	13	32	0,29	0,15	0,43	04-Neutro
PLANALTINA	0,02	0,7	4	5	0,44	0,04	0,85	04-Neutro
PONTA GROSSA	1,68	0,98	384	344	0,53	0,49	0,56	04-Neutro
PORTO ALEGRE	10,24	0,89	2.238	2.198	0,5	0,49	0,52	05-Mau
PORTO VELHO	0,36	0,18	26	130	0,17	0,11	0,23	04-Neutro
RIACHO FUNDO	0,3	0,58	51	77	0,4	0,31	0,48	04-Neutro
RIO BRANCO	0,09	0,13	5	33	0,13	0,02	0,24	04-Neutro
RIO GRANDE	0,18	1,2	45	33	0,58	0,46	0,69	04-Neutro
ROLANDIA	0,28	1,53	77	44	0,64	0,55	0,72	04-Neutro
SANTA CRUZ DO SUL	0,37	2,63	120	40	0,75	0,68	0,82	04-Neutro
SANTA MARIA	2,27	1,29	585	399	0,59	0,56	0,63	03-Bom
SAO JOSE	1,98	1,57	551	307	0,64	0,61	0,67	02-Muito bom
SAO JOSE DOS PINHAIS	0,33	0,81	69	75	0,48	0,4	0,56	04-Neutro

SAO LEOPOLDO	0,58	1,08	138	112	0,55	0,49	0,61	04-Neutro
SAPIRANGA	0,15	0,66	28	37	0,43	0,31	0,55	04-Neutro
SAPUCAIA DO SUL	0,22	0,41	30	64	0,32	0,22	0,42	04-Neutro
SARANDI	0,17	0,6	30	44	0,41	0,29	0,52	04-Neutro
SOBRADINHO	0,26	1,1	63	50	0,56	0,46	0,65	04-Neutro
TAGUATINGA	5,02	0,73	985	1.189	0,45	0,43	0,47	05-Mau
TOLEDO	0,28	2,49	91	32	0,74	0,66	0,82	04-Neutro
TUBARAO	0,07	1,95	20	9	0,69	0,51	0,87	04-Neutro
UMUARAMA	0,56	0,91	124	120	0,51	0,45	0,57	04-Neutro
VARZEA GRANDE	0,46	0,23	42	159	0,21	0,15	0,27	04-Neutro
VIAMAO	0,49	0,21	41	172	0,19	0,14	0,25	04-Neutro

CdProdVoz

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	0,16	3,21	55	15	0,79	0,69	0,88	00-Missing
1-108	0,21	2,16	64	26	0,71	0,62	0,81	04-Neutro
1-1AS	0,01	0,88	2	2	0,5	-0,42	1,42	04-Neutro
1-1OZBO3	0,04	0,7	8	10	0,44	0,19	0,7	04-Neutro
1-1OZBO0	0,02	0,58	4	6	0,4	0,03	0,77	04-Neutro
1-1OZBOA	0,4	0,9	87	85	0,51	0,43	0,58	04-Neutro
1-1OZBO6	0,12	1,31	30	20	0,6	0,46	0,74	04-Neutro
1-1OZBO7	0,04	0,37	5	12	0,29	0,05	0,54	04-Neutro
1-30VGRU	0,01	0,58	2	3	0,4	-0,28	1,08	04-Neutro
1-3CPXQ2	0	0	0	2	0	0	0	04-Neutro
1-3CPXQ3	0,02	5,26	6	1	0,86	0,51	1,21	04-Neutro
1-00VENP	2,03	3,15	687	191	0,78	0,76	0,81	01-Excelente
1-00VENQ	0,89	1,65	253	134	0,65	0,61	0,7	02-Muito bom
1-00VENR	0,02	0,88	5	5	0,5	0,12	0,88	04-Neutro
1-0KZFG0	34,43	1,85	10.129	4.792	0,68	0,67	0,69	02-Muito bom
1-0OYDNY	1,55	1,04	364	307	0,54	0,5	0,58	04-Neutro
1-A8KRG7	0,45	1,62	126	68	0,65	0,58	0,72	04-Neutro
1-A8KRG8	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro
1-A8KRG9	0,01	0,58	2	3	0,4	-0,28	1,08	04-Neutro
1-A8KRGA	0,15	1,53	42	24	0,64	0,52	0,76	04-Neutro
1-A8KRGB	0,03	1,02	7	6	0,54	0,22	0,85	04-Neutro
1-A8KRGC	0,01	1,75	4	2	0,67	0,12	1,21	04-Neutro

1-8C167L	0,2	0,54	33	54	0,38	0,28	0,48	04-Neutro
1-8C167Z	0,08	2,74	25	8	0,76	0,6	0,91	04-Neutro
1-8C167N	15,13	0,7	2.906	3.651	0,44	0,43	0,46	05-Mau
1-8C167O	6,39	0,38	830	1.939	0,3	0,28	0,32	07-Horrivel
1-8LFRJD	0,03	0,19	2	9	0,18	-0,09	0,45	04-Neutro
1-8LFRJE	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro
1-8LFRJF	0,02	0	0	9	0	0	0	04-Neutro
1-8LFRJH	0,02	0,35	2	5	0,29	-0,17	0,74	04-Neutro
1-8LFRJT	0,08	0,12	4	29	0,12	0	0,24	04-Neutro
1-8LFRJU	14,34	1,59	4.005	2.211	0,64	0,63	0,66	02-Muito bom
1-8LFRJV	10,64	0,26	1.056	3.555	0,23	0,22	0,24	07-Horrivel
1-8LFRJX	2,75	0,2	218	973	0,18	0,16	0,21	07-Horrivel
1-8LFRJY	0,9	0,13	51	340	0,13	0,1	0,16	07-Horrivel
1-8LFRJZ	0,23	0,13	13	88	0,13	0,06	0,2	04-Neutro
1-8LFRK0	0,29	0,04	5	121	0,04	0,01	0,07	04-Neutro
1-8LFRK1	0,13	0,14	8	50	0,14	0,05	0,23	04-Neutro
1-8LFRK2	0,02	0,25	2	7	0,22	-0,12	0,56	04-Neutro
1-8LFRK3	0,12	0,14	7	43	0,14	0,04	0,24	04-Neutro
1-8LFRK0	0,01	0	0	5	0	0	0	04-Neutro
1-8LFRKA	0,01	0,29	1	3	0,25	-0,55	1,05	04-Neutro
1-8LFRK6	2,27	0,89	496	488	0,5	0,47	0,54	05-Mau
1-8LFRK7	1,29	0,54	215	346	0,38	0,34	0,42	06-Muito Mau
1-8LFRK8	0,42	0,34	51	133	0,28	0,21	0,34	04-Neutro
1-8LFRK9	0,14	0,13	8	53	0,13	0,04	0,22	04-Neutro
1-QOU1P	2,62	2,96	875	259	0,77	0,75	0,8	01-Excelente
1-QOU1Q	1,04	2,25	323	126	0,72	0,68	0,76	01-Excelente
1-QOU1R	0,05	2	16	7	0,7	0,49	0,9	04-Neutro
T-ZZ	0,02	7,89	9	1	0,9	0,67	1,13	04-Neutro
T-FK	0,11	2,48	34	12	0,74	0,61	0,87	04-Neutro

DsProdVoz

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	0,16	3,21	55	15	0,79	0,69	0,88	00-Missing
Assinatura XX Adaptavel 1000	0,45	1,62	126	68	0,65	0,58	0,72	04-Neutro
Assinatura XX Adaptavel 10000	0,01	0,44	1	2	0,33	-1,1	1,77	04-Neutro

Assinatura XX Adaptavel 15000	0,01	0,58	2	3	0,4	-0,28	1,08	04-Neutro
Assinatura XX Adaptavel 2000	0,15	1,53	42	24	0,64	0,52	0,76	04-Neutro
Assinatura XX Adaptavel 4500	0,03	1,02	7	6	0,54	0,22	0,85	04-Neutro
Assinatura XX Adaptavel 7000	0,01	1,75	4	2	0,67	0,12	1,21	04-Neutro
Assinatura Mensal - Plano XX 10000	0,13	0,14	8	50	0,14	0,05	0,23	04-Neutro
Assinatura Mensal - Plano XX 12500	0,02	0,25	2	7	0,22	-0,12	0,56	04-Neutro
Assinatura Mensal - Plano XX 1500	2,75	0,2	218	973	0,18	0,16	0,21	07-Horrivel
Assinatura Mensal - Plano XX 15000	0,12	0,14	7	43	0,14	0,04	0,24	04-Neutro
Assinatura Mensal - Plano XX 20000	0,01	0	0	5	0	0	0	04-Neutro
Assinatura Mensal - Plano XX 25000	0,01	0,29	1	3	0,25	-0,55	1,05	04-Neutro
Assinatura Mensal - Plano XX 3000	0,9	0,13	51	340	0,13	0,1	0,16	07-Horrivel
Assinatura Mensal - Plano XX 5000	0,23	0,13	13	88	0,13	0,06	0,2	04-Neutro
Assinatura Mensal - Plano XX 7500	0,29	0,04	5	121	0,04	0,01	0,07	04-Neutro
Assinatura Mensal - Plano XX 800	10,64	0,26	1.056	3.555	0,23	0,22	0,24	07-Horrivel
Assinatura Mensal - Plano XX Adaptavel 1000	0,01	0,22	1	4	0,2	-0,36	0,76	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 10000	0,01	0,18	1	5	0,17	-0,26	0,6	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 1250	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 1500	1,3	0,54	215	348	0,38	0,34	0,42	06-Muito Mau
Assinatura Mensal - Plano XX Adaptavel 15000	0,02	0	0	7	0	0	0	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 2500	0,01	0,88	2	2	0,5	-0,42	1,42	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 25000	0,01	0	0	3	0	0	0	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 3000	0,42	0,34	51	133	0,28	0,21	0,34	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 5000	0,14	0,13	8	53	0,13	0,04	0,22	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 7500	0,08	0,12	4	29	0,12	0	0,24	04-Neutro
Assinatura Mensal - Plano XX Adaptavel 800	2,27	0,89	496	488	0,5	0,47	0,54	05-Mau
Assinatura Mensal - Plano SOHO Empresa 1000	1,55	1,04	364	307	0,54	0,5	0,58	04-Neutro
Assinatura Mensal - Plano SOHO Empresa 10000	0,04	0,7	8	10	0,44	0,19	0,7	04-Neutro
Assinatura Mensal - Plano SOHO Empresa 15000	0,02	0,58	4	6	0,4	0,03	0,77	04-Neutro

	0,16	3,21	55	15	0,79	0,69	0,88	00-Missing
Controle	0	0	0	2	0	0	0	04-Neutro
XX	17,27	0,29	1.859	5.625	0,25	0,24	0,26	07-Horrivel
XX Adaptável	4,94	0,71	961	1.180	0,45	0,43	0,47	05-Mau
Outros	0,35	2,37	111	41	0,73	0,66	0,8	04-Neutro
Planos Especiais	6,67	2,63	2.167	723	0,75	0,73	0,77	01-Excelente
Planos Tempo	21,8	0,59	3.794	5.652	0,4	0,39	0,41	06-Muito Mau
Transmit	0,01	0,58	2	3	0,4	-0,28	1,08	04-Neutro
Teste	48,77	1,77	14.134	7.003	0,67	0,66	0,68	02-Muito bom

SubGpProdVoz

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	0,16	3,21	55	15	0,79	0,69	0,88	00-Missing
Controle	0	0	0	2	0	0	0	04-Neutro
Eco High	1,38	0,21	114	486	0,19	0,16	0,22	07-Horrivel
Eco Low	20,82	0,38	2.706	6.319	0,3	0,29	0,31	07-Horrivel
Outros	0,35	2,37	111	41	0,73	0,66	0,8	04-Neutro
Planos Tempo	21,8	0,59	3.794	5.652	0,4	0,39	0,41	06-Muito Mau
Pln Esp High	2,01	1,93	599	272	0,69	0,66	0,72	02-Muito bom
Pln Esp Low	4,66	3,05	1.568	451	0,78	0,76	0,79	01-Excelente
Transmit	0,01	0,58	2	3	0,4	-0,28	1,08	04-Neutro
Teste	48,77	1,77	14.134	7.003	0,67	0,66	0,68	02-Muito bom

CdProdDados

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	41,96	0,54	6.921	11.264	0,38	0,37	0,39	00-Missing
23+2323F+4	0,1	12,27	42	3	0,93	0,86	1,01	04-Neutro
23#23XPYN5	0,03	3,51	12	3	0,8	0,57	1,03	04-Neutro
23#23XPYN6	0,01	4,38	5	1	0,83	0,4	1,26	04-Neutro
23#35CAFL	0,56	1,64	157	84	0,65	0,59	0,71	02-Muito bom
23#48TOSX	0,07	12,27	28	2	0,93	0,84	1,03	04-Neutro
23#4AZFGC	0,05	1,75	14	7	0,67	0,45	0,89	04-Neutro
23#4AZFGD	0,08	8,77	30	3	0,91	0,81	1,01	04-Neutro

23#57665J	0,82	1,35	215	140	0,61	0,55	0,66	03-Bom
23#6LR2AA	0,05	0,21	4	17	0,19	0,01	0,37	04-Neutro
23#6LR2AL	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
23#6LR2AM	2,18	0,63	394	552	0,42	0,39	0,45	06-Muito Mau
23#6LR2AN	19,34	1,54	5.344	3.036	0,64	0,63	0,65	02-Muito bom
23#6LR2AO	2,82	1,19	705	519	0,58	0,55	0,6	03-Bom
23#6LR2AP	16,27	1,92	4.842	2.211	0,69	0,68	0,7	02-Muito bom
23#6LR2AQ	1	0,49	156	279	0,36	0,31	0,4	07-Horrivel
23#6LR2AR	4,17	1,61	1.171	636	0,65	0,63	0,67	02-Muito bom
23#6LR2AS	0,14	0,58	24	36	0,4	0,27	0,53	04-Neutro
23#6LR2AT	0,04	1,1	10	8	0,56	0,3	0,81	04-Neutro
23#6LR2AU	1,46	1,16	360	271	0,57	0,53	0,61	03-Bom
23#6LR2AV	7,2	1,92	2.142	980	0,69	0,67	0,7	02-Muito bom
23#6LR2AW	0,27	0,75	55	64	0,46	0,37	0,55	04-Neutro
23#6LR2AX	1,28	2,95	427	127	0,77	0,74	0,81	01-Excelente

DsProdDados

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	41,96	0,54	6.921	11.264	0,38	0,37	0,39	00-Missing
Plano WebLine	0,56	1,64	157	84	0,65	0,59	0,71	02-Muito bom
Plano WebLine Teste	0,82	1,35	215	140	0,61	0,55	0,66	03-Bom
Serviço pacote 150Kbps	0,07	12,27	28	2	0,93	0,84	1,03	04-Neutro
Serviço pacote 1Mbps	0,01	4,38	5	1	0,83	0,4	1,26	04-Neutro
Serviço pacote 300Kbps	0,1	11,69	40	3	0,93	0,85	1,01	04-Neutro
Serviço pacote 300Kbps Uni	0,04	1,75	12	6	0,67	0,43	0,91	04-Neutro
Serviço pacote 300Kbps Teste	0,01	1,75	2	1	0,67	-0,77	2,1	04-Neutro
Serviço pacote 600Kbps	0,03	3,51	12	3	0,8	0,57	1,03	04-Neutro
Serviço pacote 600Kbps Uni	0,06	10,08	23	2	0,92	0,81	1,03	04-Neutro
Serviço pacote 600Kbps Teste	0,02	5,26	6	1	0,86	0,51	1,21	04-Neutro
pacote Ultra 10Mbps	0,05	0,21	4	17	0,19	0,01	0,37	04-Neutro
pacote Ultra 10Mbps Teste	0	0,88	1	1	0,5	-5,85	6,85	04-Neutro
pacote Ultra 1Mbps	2,18	0,63	394	552	0,42	0,39	0,45	06-Muito Mau
pacote Ultra 1Mbps Teste	19,34	1,54	5.344	3.036	0,64	0,63	0,65	02-Muito bom
pacote Ultra 250Kbps	2,82	1,19	705	519	0,58	0,55	0,6	03-Bom

pacote Ultra 250Kbps Teste	16,27	1,92	4.842	2.211	0,69	0,68	0,7	02-Muito bom
pacote Ultra 2Mbps	1	0,49	156	279	0,36	0,31	0,4	07-Horrivel
pacote Ultra 2Mbps Teste	4,17	1,61	1.171	636	0,65	0,63	0,67	02-Muito bom
pacote Ultra 4Mbps	0,14	0,58	24	36	0,4	0,27	0,53	04-Neutro
pacote Ultra 4Mbps Teste	0,04	1,1	10	8	0,56	0,3	0,81	04-Neutro
pacote Ultra 500Kbps	1,46	1,16	360	271	0,57	0,53	0,61	03-Bom
pacote Ultra 500Kbps Teste	7,2	1,92	2.142	980	0,69	0,67	0,7	02-Muito bom
pacote Ultra 750Kbps	0,27	0,75	55	64	0,46	0,37	0,55	04-Neutro
pacote Ultra 750Kbps Teste	1,28	2,95	427	127	0,77	0,74	0,81	01-Excelente

TpProdDados

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	41,96	0,54	6.921	11.264	0,38	0,37	0,39	00-Missing
PLANO	1,38	1,46	372	224	0,62	0,59	0,66	03-Bom
Pacote	56,66	1,58	15.802	8.756	0,64	0,64	0,65	02-Muito bom

GpProdDados

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	41,96	0,54	6.921	11.264	0,38	0,37	0,39	00-Missing
Pacote	0,43	7,7	167	19	0,9	0,85	0,94	04-Neutro
Pacote Ultra	56,24	1,57	15.635	8.737	0,64	0,64	0,65	02-Muito bom
Webline	1,38	1,46	372	224	0,62	0,59	0,66	03-Bom

SubGpProdDados

Categoria	% Total	RR	Bom	Mau	% Bons	IC 95% Inf.	IC 95% Sup.	Classe
	41,96	0,54	6.921	11.264	0,38	0,37	0,39	00-Missing
TN High	0,13	6,01	48	7	0,87	0,78	0,96	04-Neutro
TN Low	0,3	8,69	119	12	0,91	0,86	0,96	04-Neutro
TN Ultra High	26,92	1,36	7.104	4.565	0,61	0,6	0,62	03-Bom
TN Ultra Low	29,31	1,79	8.531	4.172	0,67	0,66	0,68	02-Muito bom
Webline	1,38	1,46	372	224	0,62	0,59	0,66	03-Bom

Apêndice 2 – Análise Bivariada

CEP								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	6	1	0,03	0,00	85,71	0,17	0,15	-1,92
01-Excelente	279	741	1,38	3,21	27,35	2,66	2,33	0,85
02-Muito Bom	2.180	4.054	10,77	17,55	34,97	1,86	1,63	0,49
03-Bom	4.915	7.165	24,28	31,02	40,69	1,46	1,28	0,25
04-Neutro	3.002	3.499	14,83	15,15	46,18	1,17	1,02	0,02
05-Mau	4.911	4.448	24,26	19,26	52,47	0,91	0,79	-0,23
06-Muito mau	3.233	2.319	15,97	10,04	58,23	0,72	0,63	-0,46
07-Horrível	1.718	868	8,49	3,76	66,43	0,51	0,44	-0,81

Information Value = 14,3271

NAT_OCUP								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	269	495	1,33	2,14	35,21	1,84	1,61	0,48
02-Muito Bom	606	1.161	2,99	5,03	34,30	1,92	1,68	0,52
03-Bom	2.575	3.470	12,72	15,02	42,60	1,35	1,18	0,17
04-Neutro	8.057	9.103	39,80	39,42	46,95	1,13	0,99	-0,01
05-Mau	8.737	8.866	43,16	38,39	49,63	1,01	0,89	-0,12

Information Value = 2,3897

OCUPACAO								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	9.904	9.966	48,92	43,15	49,84	1,01	0,88	-0,13
01-Excelente	539	1.586	2,66	6,87	25,36	2,94	2,58	0,95
02-Muito Bom	515	1.057	2,54	4,58	32,76	2,05	1,80	0,59
03-Bom	1.914	2.962	9,45	12,83	39,25	1,55	1,36	0,30
04-Neutro	5.251	5.923	25,94	25,65	46,99	1,13	0,99	-0,01
05-Mau	1.116	999	5,51	4,33	52,77	0,90	0,78	-0,24
06-Muito mau	843	538	4,16	2,33	61,04	0,64	0,56	-0,58
07-Horrível	162	64	0,80	0,28	71,68	0,40	0,35	-1,06

Information Value = 8,8416

TpResidencia								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	8.720	8.864	43,07	38,38	49,59	1,02	0,89	-0,12
04-Neutro	10.322	13.027	50,99	56,41	44,21	1,26	1,11	0,10
05-Mau	1.202	1.204	5,94	5,21	49,96	1,00	0,88	-0,13

Information Value = 1,1830

EstCivil								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	7.772	8.022	38,39	34,73	49,21	1,03	0,90	-0,10
02-Muito Bom	178	317	0,88	1,37	35,96	1,78	1,56	0,45
03-Bom	360	458	1,78	1,98	44,01	1,27	1,12	0,11
04-Neutro	9.392	11.799	46,39	51,09	44,32	1,26	1,10	0,10
05-Mau	2.542	2.499	12,56	10,82	50,43	0,98	0,86	-0,15

Information Value = 1,3191

UF								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	0	1	0,00	0,00	0,00			
02-Muito Bom	279	635	1,38	2,75	30,53	2,28	2,00	0,69
03-Bom	8.138	11.019	40,20	47,71	42,48	1,35	1,19	0,17
04-Neutro	7.836	8.275	38,71	35,83	48,64	1,06	0,93	-0,08
05-Mau	3.049	2.599	15,06	11,25	53,98	0,85	0,75	-0,29
06-Muito mau	942	566	4,65	2,45	62,47	0,60	0,53	-0,64

Information Value = 4,9783

Cidade								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	0	1	0,00	0,00	0,00			
01-Excelente	272	666	1,34	2,88	29,00	2,45	2,15	0,76
02-Muito Bom	1.941	3.653	9,59	15,82	34,70	1,88	1,65	0,50
03-Bom	4.746	6.911	23,44	29,92	40,71	1,46	1,28	0,24
04-Neutro	3.622	3.565	17,89	15,44	50,40	0,98	0,86	-0,15
05-Mau	7.833	7.234	38,69	31,32	51,99	0,92	0,81	-0,21
06-Muito mau	1.065	689	5,26	2,98	60,72	0,65	0,57	-0,57
07-Horrível	765	376	3,78	1,63	67,05	0,49	0,43	-0,84

Information Value = 10,8989

CdProdVoz								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	15	55	0,07	0,24	21,43	3,67	3,21	1,17
01-Excelente	576	1.885	2,85	8,16	23,41	3,27	2,87	1,05
02-Muito Bom	7.137	14.387	35,25	62,29	33,16	2,02	1,77	0,57
04-Neutro	1.224	996	6,05	4,31	55,14	0,81	0,71	-0,34

05-Mau	4.139	3.402	20,45	14,73	54,89	0,82	0,72	-0,33
06-Muito mau	346	215	1,71	0,93	61,68	0,62	0,54	-0,61
07-Horrível	6.807	2.155	33,62	9,33	75,95	0,32	0,28	-1,28

Information Value = 55,2624

DsProdVoz								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	15	55	0,07	0,24	21,43	3,67	3,21	1,17
01-Excelente	576	1.885	2,85	8,16	23,41	3,27	2,87	1,05
02-Muito Bom	7.137	14.387	35,25	62,29	33,16	2,02	1,77	0,57
04-Neutro	1.222	996	6,04	4,31	55,09	0,82	0,71	-0,34
05-Mau	4.139	3.402	20,45	14,73	54,89	0,82	0,72	-0,33
06-Muito mau	348	215	1,72	0,93	61,81	0,62	0,54	-0,61
07-Horrível	6.807	2.155	33,62	9,33	75,95	0,32	0,28	-1,28

Information Value = 55,2668

TpProdVoz								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	15	55	0,07	0,24	21,43	3,67	3,21	1,17
03-Bom	13.424	20.220	66,31	87,55	39,90	1,51	1,32	0,28
07-Horrível	6.805	2.820	33,61	12,21	70,70	0,41	0,36	-1,01

Information Value = 27,7695

GpProdVoz								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	15	55	0,07	0,24	21,43	3,67	3,21	1,17
01-Excelente	723	2.167	3,57	9,38	25,02	3,00	2,63	0,97
02-Muito Bom	7.003	14.134	34,59	61,20	33,13	2,02	1,77	0,57
04-Neutro	46	125	0,23	0,54	26,90	2,72	2,38	0,87
05-Mau	1.180	961	5,83	4,16	55,11	0,81	0,71	-0,34
06-Muito mau	5.652	3.794	27,92	16,43	59,83	0,67	0,59	-0,53
07-Horrível	5.625	1.859	27,79	8,05	75,16	0,33	0,29	-1,24

Information Value = 52,3654

SubGpProdVoz								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	15	55	0,07	0,24	21,43	3,67	3,21	1,17
01-Excelente	451	1.568	2,23	6,79	22,34	3,48	3,05	1,11
02-Muito Bom	7.275	14.733	35,94	63,79	33,06	2,03	1,78	0,57

04-Neutro	46	125	0,23	0,54	26,90	2,72	2,38	0,87
06-Muito mau	5.652	3.794	27,92	16,43	59,83	0,67	0,59	-0,53
07-Horrível	6.805	2.820	33,61	12,21	70,70	0,41	0,36	-1,01

Information Value = 49,3040

CdProdDados								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WO E
00-Missing	11.264	6.921	55,64	29,97	61,94	0,61	0,54	-0,62
01-Excelente	127	427	0,63	1,85	22,92	3,36	2,95	1,08
02-Muito Bom	6.947	13.656	34,32	59,13	33,72	1,97	1,72	0,54
03-Bom	930	1.280	4,59	5,54	42,08	1,38	1,21	0,19
04-Neutro	145	261	0,72	1,13	35,71	1,80	1,58	0,46
06-Muito mau	552	394	2,73	1,71	58,35	0,71	0,63	-0,47
07-Horrível	279	156	1,38	0,68	64,14	0,56	0,49	-0,71

Information Value = 32,0551

DsProdDados								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WO E
00-Missing	11.264	6.921	55,64	29,97	61,94	0,61	0,54	-0,62
01-Excelente	127	427	0,63	1,85	22,92	3,36	2,95	1,08
02-Muito Bom	6.947	13.656	34,32	59,13	33,72	1,97	1,72	0,54
03-Bom	930	1.280	4,59	5,54	42,08	1,38	1,21	0,19
04-Neutro	145	261	0,72	1,13	35,71	1,80	1,58	0,46
06-Muito mau	552	394	2,73	1,71	58,35	0,71	0,63	-0,47
07-Horrível	279	156	1,38	0,68	64,14	0,56	0,49	-0,71

Information Value = 32,0551

TpProdDados								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WO E
00-Missing	11.264	6.921	55,64	29,97	61,94	0,61	0,54	-0,62
02-Muito Bom	8.756	15.802	43,25	68,42	35,65	1,80	1,58	0,46
03-Bom	224	372	1,11	1,61	37,58	1,66	1,46	0,38

Information Value = 27,6201

GpProdDados								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WO E
00-Missing	11.264	6.921	55,64	29,97	61,94	0,61	0,54	-0,62
02-Muito Bom	8.737	15.635	43,16	67,70	35,85	1,79	1,57	0,45
03-Bom	224	372	1,11	1,61	37,58	1,66	1,46	0,38

04-Neutro	19	167	0,09	0,72	10,22	8,79	7,70	2,04
-----------	----	-----	------	------	-------	------	-------------	-------------

Information Value = 28,4089

SubGpProdDados								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	11.264	6.921	55,64	29,97	61,94	0,61	0,54	-0,62
02-Muito Bom	4.172	8.531	20,61	36,94	32,84	2,04	1,79	0,58
03-Bom	4.789	7.476	23,66	32,37	39,05	1,56	1,37	0,31
04-Neutro	19	167	0,09	0,72	10,22	8,79	7,70	2,04

Information Value = 29,4344

Parcelamento								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
1) Yes	13.233	57	65,37	0,25	99,57	0,00	0,00	-5,58
2) No	7.011	23.038	34,63	99,75	23,33	3,29	2,88	1,06

Information Value = 432,2115

Internet								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
1) Yes	9.135	16.227	45,12	70,26	36,02	1,78	1,56	0,44
2) No	11.109	6.868	54,88	29,74	61,80	0,62	0,54	-0,61

Information Value = 26,5313

FlagFraude								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
1) Yes	18.274	22.771	90,27	98,60	44,52	1,25	1,09	0,09
2) No	1.970	324	9,73	1,40	85,88	0,16	0,14	-1,94

Information Value = 16,8654

FlagDesc								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
1) Yes	19.200	20.654	94,84	89,43	48,18	1,08	0,94	-0,06
2) No	1.044	2.441	5,16	10,57	29,96	2,34	2,05	0,72

Information Value = 4,2018

Migracao								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
0) Missing	173	930	0,85	4,03	15,68	5,38	4,71	1,55

1) Yes	8.529	11.531	42,13	49,93	42,52	1,35	1,19	0,17
2) No	11.542	10.634	57,01	46,04	52,05	0,92	0,81	-0,21

Information Value = 8,5857

StatusInst								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
Ativo parcial	11.070	6.098	54,68	26,40	64,48	0,55	0,48	-0,73
Ativo total	9.174	16.997	45,32	73,60	35,05	1,85	1,62	0,48

Information Value = 34,3006

TipoTecnologia								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
ANTENA	5.001	2.808	24,70	12,16	64,04	0,56	0,49	-0,71
CABO	15.243	20.287	75,30	87,84	42,90	1,33	1,17	0,15

Information Value = 10,8268

CanalVendas								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00-Missing	29	257	0,14	1,11	10,14	8,86	7,77	2,05
BCENT	11.411	11.514	56,37	49,85	49,78	1,01	0,88	-0,12
CALL	262	832	1,29	3,60	23,95	3,18	2,78	1,02
DEAL	6.954	8.239	34,35	35,67	45,77	1,18	1,04	0,04
PORT	163	226	0,81	0,98	41,90	1,39	1,22	0,20
SVEND	1.425	2.027	7,04	8,78	41,28	1,42	1,25	0,22

Information Value = 5,6174

WhiteList								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
0	71	828	0,35	3,59	7,90	11,66	10,22	2,32
00-Missing	12.599	7.362	62,24	31,88	63,12	0,58	0,51	-0,67
1	285	2.363	1,41	10,23	10,76	8,29	7,27	1,98
2	466	2.816	2,30	12,19	14,20	6,04	5,30	1,67
3	789	2.663	3,90	11,53	22,86	3,38	2,96	1,08
4	953	2.127	4,71	9,21	30,94	2,23	1,96	0,67
5	1.221	1.806	6,03	7,82	40,34	1,48	1,30	0,26
6	1.460	1.270	7,21	5,50	53,48	0,87	0,76	-0,27
7	969	862	4,79	3,73	52,92	0,89	0,78	-0,25
8	804	603	3,97	2,61	57,14	0,75	0,66	-0,42
9	588	357	2,90	1,55	62,22	0,61	0,53	-0,63

B	6	11	0,03	0,05	35,29	1,83	1,61	0,47
C	2	6	0,01	0,03	25,00	3,00	2,63	0,97
D	2	1	0,01	0,00	66,67	0,50	0,44	-0,82
J	28	17	0,14	0,07	62,22	0,61	0,53	-0,63
N	1	3	0,00	0,01	25,00	3,00	2,63	0,97

Information Value = 75,8187

Categoria								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
Residencial	19.579	22.626	96,72	97,97	46,39	1,16	1,01	0,01
Retail	665	469	3,28	2,03	58,64	0,71	0,62	-0,48

Information Value = 0,6194

VIFatura								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 20,62	647	1.518	3,20	6,57	29,88	2,35	2,06	0,72
02. > 20,62 e ≤ 82,25	7.213	3.624	35,63	15,69	66,56	0,50	0,44	-0,82
03. > 82,25 e ≤ 92,19	1.267	897	6,26	3,88	58,55	0,71	0,62	-0,48
04. > 92,19 e ≤ 111,33	2.177	2.157	10,75	9,34	50,23	0,99	0,87	-0,14
05. > 111,33 e ≤ 132,38	1.944	2.392	9,60	10,36	44,83	1,23	1,08	0,08
06. > 132,38 e ≤ 171,59	2.564	3.937	12,67	17,05	39,44	1,54	1,35	0,30
07. > 171,59	4.432	8.570	21,89	37,11	34,09	1,93	1,69	0,53

Information Value = 29,5051

VIAjuste								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 1,49	18.242	20.724	90,11	89,73	46,82	1,14	1,00	0,00
02. > 1,49 e ≤ 5,28	772	1.434	3,81	6,21	35,00	1,86	1,63	0,49
03. > 5,28	1.230	937	6,08	4,06	56,76	0,76	0,67	-0,40

Information Value = 1,9847

TotalPago								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 1,90	3.954	378	19,53	1,64	91,27	0,10	0,08	-2,48
02. > 1,90 e ≤ 40,40	6.561	2.106	32,41	9,12	75,70	0,32	0,28	-1,27
03. > 40,40 e ≤ 51,89	1.497	674	7,39	2,92	68,95	0,45	0,39	-0,93
04. > 51,89 e ≤ 76,15	2.584	1.747	12,76	7,56	59,66	0,68	0,59	-0,52
05. > 76,15 e ≤ 88,74	1.027	1.141	5,07	4,94	47,37	1,11	0,97	-0,03
06. > 88,74 e ≤ 100,71	811	1.357	4,01	5,88	37,41	1,67	1,47	0,38

07. > 100,71 e ≤ 141,30	1.743	4.757	8,61	20,60	26,82	2,73	2,39	0,87
08. > 141,30 e ≤ 178,64	803	3.533	3,97	15,30	18,52	4,40	3,86	1,35
09. > 178,64	1.264	7.402	6,24	32,05	14,59	5,86	5,13	1,64

Information Value = 149,4678

VIFaturaAj								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	1.322	16.147	6,53	69,92	7,57	12,21	10,71	2,37
02. > 0,00 e ≤ 19,45	2.800	5.757	13,83	24,93	32,72	2,06	1,80	0,59
03. > 19,45 e ≤ 34,65	3.325	1.005	16,42	4,35	76,79	0,30	0,26	-1,33
04. > 34,65 e ≤ 55,07	4.172	156	20,61	0,68	96,40	0,04	0,03	-3,42
05. > 55,07	8.625	30	42,61	0,13	99,65	0,00	0,00	-5,79

487,038

Information Value = 8

MediaPago								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00. Missing	2.180	148	10,77	0,64	93,64	0,07	0,06	-2,82
01. ≤ 15,00	2.016	316	9,96	1,37	86,45	0,16	0,14	-1,98
02. > 15,00 e ≤ 31,55	1.246	522	6,15	2,26	70,48	0,42	0,37	-1,00
03. > 31,55 e ≤ 83,19	7.284	5.019	35,98	21,73	59,21	0,69	0,60	-0,50
04. > 83,19 e ≤ 90,71	949	1.104	4,69	4,78	46,23	1,16	1,02	0,02
05. > 90,71 e ≤ 97,48	761	1.288	3,76	5,58	37,14	1,69	1,48	0,39
06. > 97,48 e ≤ 110,30	1.242	2.861	6,14	12,39	30,27	2,30	2,02	0,70
07. > 110,30 e ≤ 222,56	3.646	10.707	18,01	46,36	25,40	2,94	2,57	0,95
08. > 222,56	920	1.130	4,54	4,89	44,88	1,23	1,08	0,07

Information Value = 88,6575

MaiorVIPago								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
00. Missing	2.180	148	10,77	0,64	93,64	0,07	0,06	-2,82
01. ≤ 15,00	1.975	282	9,76	1,22	87,51	0,14	0,13	-2,08
02. > 15,00 e ≤ 47,21	1.204	640	5,95	2,77	65,29	0,53	0,47	-0,76
03. > 47,21 e ≤ 101,91	3.172	2.979	15,67	12,90	51,57	0,94	0,82	-0,19
04. > 101,91 e ≤ 123,78	1.634	2.468	8,07	10,69	39,83	1,51	1,32	0,28
05. > 123,78 e ≤ 242,26	6.304	12.151	31,14	52,61	34,16	1,93	1,69	0,52
06. > 242,26	3.775	4.427	18,65	19,17	46,03	1,17	1,03	0,03

Information Value = 61,2885

QtdeLinhas

ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 1	10.223	6.060	50,50	26,24	62,78	0,59	0,52	-0,65
02. > 1	10.021	17.035	49,50	73,76	37,04	1,70	1,49	0,40

Information Value = 25,5578

VIFatAtrasoMax								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 145,08	9.284	16.720	45,86	72,40	35,70	1,80	1,58	0,46
02. > 145,08 e ≤ 172,02	1.861	2.473	9,19	10,71	42,94	1,33	1,16	0,15
03. > 172,02 e ≤ 214,84	2.359	1.974	11,65	8,55	54,44	0,84	0,73	-0,31
04. > 214,84 e ≤ 250,43	1.412	755	6,97	3,27	65,16	0,53	0,47	-0,76
05. > 250,43 e ≤ 313,86	1.559	608	7,70	2,63	71,94	0,39	0,34	-1,07
06. > 313,86 e ≤ 455,93	1.779	388	8,79	1,68	82,10	0,22	0,19	-1,65
07. > 455,93	1.990	177	9,83	0,77	91,83	0,09	0,08	-2,55

Information Value = 56,4437

SeqFatura								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 9	14.535	3.929	71,80	17,01	78,72	0,27	0,24	-1,44
02. > 9 e ≤ 10	1.825	1.766	9,02	7,65	50,82	0,97	0,85	-0,16
03. > 10 e ≤ 12	2.493	5.179	12,31	22,42	32,49	2,08	1,82	0,60
04. > 12 e ≤ 14	925	3.487	4,57	15,10	20,97	3,77	3,30	1,20
05. > 14 e ≤ 17	384	4.179	1,90	18,09	8,42	10,88	9,54	2,26
06. > 17	82	4.555	0,41	19,72	1,77	55,55	48,69	3,89

Information Value = 209,3519

L1feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 96,55	8.454	6.713	41,76	29,07	55,74	0,79	0,70	-0,36
02. > 96,55 e ≤ 163,95	4.199	4.470	20,74	19,35	48,44	1,06	0,93	-0,07
03. > 163,95 e ≤ 236,58	2.772	3.730	13,69	16,15	42,63	1,35	1,18	0,17
04. > 236,58	4.819	8.182	23,80	35,43	37,07	1,70	1,49	0,40

Information Value = 9,7227

L2feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	16.197	18.021	80,01	78,03	47,33	1,11	0,98	-0,03
02. > 0,00 e ≤ 44,80	1.768	3.019	8,73	13,07	36,93	1,71	1,50	0,40
03. > 44,80 e ≤ 516,14	972	1.195	4,80	5,17	44,85	1,23	1,08	0,07

04. > 516,14	1.307	860	6,46	3,72	60,31	0,66	0,58	-0,55
--------------	-------	-----	------	------	-------	------	-------------	--------------

Information Value = 3,3310

L3feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.243	23.079	100,00	99,93	46,73	1,14	1,00	0,00
02. >= 0,00	1	16	0,00	0,07	5,88	16,00	14,02	2,64

Information Value = 0,1700

L4feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.234	23.022	99,95	99,68	46,78	1,14	1,00	0,00
02. >= 0,00	10	73	0,05	0,32	12,05	7,30	6,40	1,86

Information Value = 0,4957

L5feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.235	23.023	99,96	99,69	46,78	1,14	1,00	0,00
02. >= 0,00	9	72	0,04	0,31	11,11	8,00	7,01	1,95

Information Value = 0,5213

L6feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	16.925	18.476	83,61	80,00	47,81	1,09	0,96	-0,04
02. >= 0,00	3.319	4.619	16,39	20,00	41,81	1,39	1,22	0,20

Information Value = 0,8754

L7feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 3,37	2.170	6.498	10,72	28,14	25,03	2,99	2,62	0,97
02. > 3,37 e ≤ 6,69	1.249	3.081	6,17	13,34	28,85	2,47	2,16	0,77
03. > 6,69 e ≤ 10,91	1.494	2.844	7,38	12,31	34,44	1,90	1,67	0,51
04. > 10,91 e ≤ 16,12	1.771	2.564	8,75	11,10	40,85	1,45	1,27	0,24
05. > 16,12 e ≤ 27,72	3.098	3.399	15,30	14,72	47,68	1,10	0,96	-0,04
06. > 27,72 e ≤ 33,52	1.149	1.019	5,68	4,41	53,00	0,89	0,78	-0,25
07. > 33,52 e ≤ 68,46	3.993	2.510	19,72	10,87	61,40	0,63	0,55	-0,60
08. > 68,46 e ≤ 100,67	1.538	628	7,60	2,72	71,01	0,41	0,36	-1,03
09. > 100,67 e ≤ 186,54	1.775	392	8,77	1,70	81,91	0,22	0,19	-1,64
10. > 186,54	2.007	160	9,91	0,69	92,62	0,08	0,07	-2,66

Information Value = 72,2039

L8feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	9.058	11.098	44,74	48,05	44,94	1,23	1,07	0,07
02. > 0,00 e ≤ 3,56	3.908	6.277	19,30	27,18	38,37	1,61	1,41	0,34
03. > 3,56 e ≤ 9,85	2.853	3.646	14,09	15,79	43,90	1,28	1,12	0,11
04. > 9,85 e ≤ 15,83	1.191	974	5,88	4,22	55,01	0,82	0,72	-0,33
05. > 15,83 e ≤ 31,89	1.434	733	7,08	3,17	66,17	0,51	0,45	-0,80
06. > 31,89	1.800	367	8,89	1,59	83,06	0,20	0,18	-1,72

Information Value = 19,3902

L9feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 8,73	7.008	5.989	34,62	25,93	53,92	0,85	0,75	-0,29
02. > 8,73 e ≤ 12,02	2.020	2.319	9,98	10,04	46,55	1,15	1,01	0,01
03. > 12,02 e ≤ 29,96	7.190	10.144	35,52	43,92	41,48	1,41	1,24	0,21
04. > 29,96 e ≤ 52,86	2.899	3.603	14,32	15,60	44,59	1,24	1,09	0,09
05. > 52,86	1.127	1.040	5,57	4,50	52,01	0,92	0,81	-0,21

Information Value = 4,6306

L10feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 2,78	8.463	11.040	41,80	47,80	43,39	1,30	1,14	0,13
02. > 2,78 e ≤ 5,99	3.940	4.722	19,46	20,45	45,49	1,20	1,05	0,05
03. > 5,99	7.841	7.333	38,73	31,75	51,67	0,94	0,82	-0,20

Information Value = 2,2399

L11feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 0,62	3.757	2.735	18,56	11,84	57,87	0,73	0,64	-0,45
02. > 0,62 e ≤ 1,85	1.157	1.016	5,72	4,40	53,24	0,88	0,77	-0,26
03. > 1,85 e ≤ 3,78	2.037	2.306	10,06	9,98	46,90	1,13	0,99	-0,01
04. > 3,78 e ≤ 14,37	8.126	11.376	40,14	49,26	41,67	1,40	1,23	0,20
05. > 14,37 e ≤ 20,27	1.913	2.419	9,45	10,47	44,16	1,26	1,11	0,10
06. > 20,27	3.254	3.243	16,07	14,04	50,08	1,00	0,87	-0,14

Information Value = 5,6084

L12feitos								
------------------	--	--	--	--	--	--	--	--

ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.238	23.029	99,97	99,71	46,77	1,14	1,00	0,00
02. >= 0,00	6	66	0,03	0,29	8,33	11,00	9,64	2,27

Information Value = 0,5811

L15feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 0,40	12.641	15.520	62,44	67,20	44,89	1,23	1,08	0,07
02. > 0,40 e ≤ 5,55	3.624	5.051	17,90	21,87	41,78	1,39	1,22	0,20
03. > 5,55 e ≤ 9,51	1.036	1.134	5,12	4,91	47,74	1,09	0,96	-0,04
04. > 9,51 e ≤ 20,41	1.263	903	6,24	3,91	58,31	0,71	0,63	-0,47
05. > 20,41	1.680	487	8,30	2,11	77,53	0,29	0,25	-1,37

Information Value = 10,7216

L16feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	19.973	22.631	98,66	97,99	46,88	1,13	0,99	-0,01
02. >= 0,00	271	464	1,34	2,01	36,87	1,71	1,50	0,41

Information Value = 0,2768

L17feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.244	23.094	100,00	100,00	46,71	1,14	1,00	0,00
02. >= 0,00	0	1	0,00	0,00	0,00			

Information Value = 0,0000

L18feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	11.020	10.372	54,44	44,91	51,51	0,94	0,83	-0,19
02. > 0,00 e ≤ 1,59	2.056	2.548	10,16	11,03	44,66	1,24	1,09	0,08
03. > 1,59 e ≤ 7,94	3.422	5.251	16,90	22,74	39,46	1,53	1,35	0,30
04. > 7,94 e ≤ 35,56	2.744	3.759	13,55	16,28	42,20	1,37	1,20	0,18
05. > 35,56	1.002	1.165	4,95	5,04	46,24	1,16	1,02	0,02

Information Value = 4,1338

L19feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 0,48	12.532	11.312	61,90	48,98	52,56	0,90	0,79	-0,23

02. > 0,48 e ≤ 26,69	6.107	9.054	30,17	39,20	40,28	1,48	1,30	0,26
03. > 26,69	1.605	2.729	7,93	11,82	37,03	1,70	1,49	0,40

Information Value = 6,9459

L20feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	11.058	9.635	54,62	41,72	53,44	0,87	0,76	-0,27
02. > 0,00 e ≤ 1,68	2.500	2.809	12,35	12,16	47,09	1,12	0,98	-0,02
03. > 1,68	6.686	10.651	33,03	46,12	38,56	1,59	1,40	0,33

Information Value = 7,8516

L21feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 0,32	9.056	8.290	44,73	35,90	52,21	0,92	0,80	-0,22
02. > 0,32 e ≤ 0,78	1.053	1.108	5,20	4,80	48,73	1,05	0,92	-0,08
03. > 0,78 e ≤ 40,18	8.369	11.130	41,34	48,19	42,92	1,33	1,17	0,15
04. > 40,18	1.766	2.567	8,72	11,11	40,76	1,45	1,27	0,24

Information Value = 3,6085

L22feitos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.237	23.043	99,97	99,77	46,76	1,14	1,00	0,00
02. ≥ 0,00	7	52	0,03	0,23	11,86	7,43	6,51	1,87

Information Value = 0,3574

L1recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 166,70	15.085	15.253	74,52	66,04	49,72	1,01	0,89	-0,12
02. > 166,70 e ≤ 225,17	1.858	2.475	9,18	10,72	42,88	1,33	1,17	0,15
03. > 225,17	3.301	5.367	16,31	23,24	38,08	1,63	1,43	0,35

Information Value = 3,7170

L2recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.235	23.087	99,96	99,97	46,71	1,14	1,00	0,00
02. ≥ 0,00	9	8	0,04	0,03	52,94	0,89	0,78	-0,25

Information Value = 0,0025

L3recebidos								
--------------------	--	--	--	--	--	--	--	--

ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.243	23.080	100,00	99,94	46,73	1,14	1,00	0,00
02. >= 0,00	1	15	0,00	0,06	6,25	15,00	13,15	2,58

Information Value = 0,1546

L4recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.229	23.016	99,93	99,66	46,78	1,14	1,00	0,00
02. >= 0,00	15	79	0,07	0,34	15,96	5,27	4,62	1,53

Information Value = 0,4106

L5recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.232	23.018	99,94	99,67	46,78	1,14	1,00	0,00
02. >= 0,00	12	77	0,06	0,33	13,48	6,42	5,62	1,73

Information Value = 0,4742

L6recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	14.756	13.414	72,89	58,08	52,38	0,91	0,80	-0,23
02. >= 0,00	5.488	9.681	27,11	41,92	36,18	1,76	1,55	0,44

Information Value = 9,8176

L7recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 5,82	5.117	7.890	25,28	34,16	39,34	1,54	1,35	0,30
02. > 5,82 e ≤ 10,08	2.782	3.720	13,74	16,11	42,79	1,34	1,17	0,16
03. > 10,08 e ≤ 15,97	2.980	3.511	14,72	15,20	45,91	1,18	1,03	0,03
04. > 15,97 e ≤ 30,95	4.256	4.413	21,02	19,11	49,09	1,04	0,91	-0,10
05. > 30,95 e ≤ 50,72	2.282	2.054	11,27	8,89	52,63	0,90	0,79	-0,24
06. > 50,72 e ≤ 78,35	1.255	912	6,20	3,95	57,91	0,73	0,64	-0,45
07. > 78,35	1.572	595	7,77	2,58	72,54	0,38	0,33	-1,10

Information Value = 10,5551

L8recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	16.375	17.296	80,89	74,89	48,63	1,06	0,93	-0,08
02. > 0,00 e ≤ 3,30	2.878	4.625	14,22	20,03	38,36	1,61	1,41	0,34

03. > 3,30	991	1.174	4,90	5,08	45,77	1,18	1,04	0,04
------------	-----	-------	------	------	-------	------	------	------

Information Value = 2,4596

L12recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.231	23.012	99,94	99,64	46,78	1,14	1,00	0,00
02. >= 0,00	13	83	0,06	0,36	13,54	6,38	5,60	1,72

Information Value = 0,5092

L15recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	18.704	20.457	92,39	88,58	47,76	1,09	0,96	-0,04
02. >= 0,00	1.540	2.638	7,61	11,42	36,86	1,71	1,50	0,41

Information Value = 1,7117

L16recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.226	23.036	99,91	99,74	46,75	1,14	1,00	0,00
02. >= 0,00	18	59	0,09	0,26	23,38	3,28	2,87	1,06

Information Value = 0,1761

L17recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.243	23.082	100,00	99,94	46,72	1,14	1,00	0,00
02. >= 0,00	1	13	0,00	0,06	7,14	13,00	11,40	2,43

Information Value = 0,1250

L18recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	9.577	8.675	47,31	37,56	52,47	0,91	0,79	-0,23
02. > 0,00 e ≤ 4,42	6.232	8.022	30,78	34,73	43,72	1,29	1,13	0,12
03. > 4,42	4.435	6.398	21,91	27,70	40,94	1,44	1,26	0,23

Information Value = 4,0852

L19recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	10.008	8.495	49,44	36,78	54,09	0,85	0,74	-0,30
02. > 0,00 e ≤ 19,33	8.848	11.654	43,71	50,46	43,16	1,32	1,15	0,14

03. > 19,33 e ≤ 37,27	758	1.409	3,74	6,10	34,98	1,86	1,63	0,49
04. > 37,27	630	1.537	3,11	6,66	29,07	2,44	2,14	0,76

Information Value = 8,5556

L20recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	3.459	3.816	17,09	16,52	47,55	1,10	0,97	-0,03
02. > 0,00 e ≤ 0,87	2.134	3.598	10,54	15,58	37,23	1,69	1,48	0,39
03. > 0,87 e ≤ 2,16	2.889	3.605	14,27	15,61	44,49	1,25	1,09	0,09
04. > 2,16 e ≤ 7,35	7.202	5.800	35,58	25,11	55,39	0,81	0,71	-0,35
05. > 7,35 e ≤ 12,98	2.092	2.244	10,33	9,72	48,25	1,07	0,94	-0,06
06. > 12,98	2.468	4.032	12,19	17,46	37,97	1,63	1,43	0,36

Information Value = 7,6798

L21recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 0,53	4.967	3.728	24,54	16,14	57,12	0,75	0,66	-0,42
02. > 0,53 e ≤ 1,34	2.094	2.201	10,34	9,53	48,75	1,05	0,92	-0,08
03. > 1,34 e ≤ 20,02	10.816	13.031	53,43	56,42	45,36	1,20	1,06	0,05
04. > 20,02 e ≤ 58,59	1.650	2.685	8,15	11,63	38,06	1,63	1,43	0,36
05. > 58,59	717	1.450	3,54	6,28	33,09	2,02	1,77	0,57

Information Value = 6,5454

L22recebidos								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. Equal to zero	20.140	22.768	99,49	98,58	46,94	1,13	0,99	-0,01
02. >= 0,00	104	327	0,51	1,42	24,13	3,14	2,76	1,01

Information Value = 0,9228

Feitos_total								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE
01. ≤ 144,94	4.643	4.024	22,94	17,42	53,57	0,87	0,76	-0,27
02. > 144,94 e ≤ 270,72	4.977	5.858	24,59	25,36	45,93	1,18	1,03	0,03
03. > 270,72 e ≤ 732,55	6.925	10.411	34,21	45,08	39,95	1,50	1,32	0,28
04. > 732,55	3.699	2.802	18,27	12,13	56,90	0,76	0,66	-0,41

Information Value = 7,0534

Recebidos_total								
ATRIBUTO	Maus	Bons	% Maus	% Bons	Taxa de Maus	Odds	RR	WOE

01. $\leq 137,37$	10.041	9.460	49,60	40,96	51,49	0,94	0,83	-0,19
02. $> 137,37$ e $\leq 249,26$	4.914	5.923	24,27	25,65	45,34	1,21	1,06	0,05
03. $> 249,26$	5.289	7.712	26,13	33,39	40,68	1,46	1,28	0,25

Information Value = 3,5117

Apêndice 3 – Agrupamento final das classes de todas as variáveis preditoras

Atributo	Information Value	% Total	Taxa de Maus	Odds	RR	WOE
Parcelamento						
1) Sim		30,67	99,57	0,00	0,00	-5,58
2) Não		69,33	23,33	3,29	2,88	1,06
Total	432,212	100,00				
Internet						
1) Sim		58,52	36,02	1,78	1,56	0,44
2) Não		41,48	61,80	0,62	0,54	-0,61
Total	26,531	100,00				
FlagFraude						
1) Não		94,71	44,52	1,25	1,09	0,09
2) Sim		5,29	85,88	0,16	0,14	-1,94
Total	16,865	100,00				
FlagDesc						
1) Não		91,96	48,18	1,08	0,94	-0,06
2) Sim		8,04	29,96	2,34	2,05	0,72
Total	4,202	100,00				
StatusInst						
Ativo parcial		39,61	64,48	0,55	0,48	-0,73
Ativo total		60,39	35,05	1,85	1,62	0,48
Total	34,301	100,00				
TipoTecnologia						
ANTENA		18,02	64,04	0,56	0,49	-0,71
CABO		81,98	42,90	1,33	1,17	0,15
Total	10,827	100,00				
Categoria						
Residencial		97,38	46,39	1,16	1,01	0,01
Retail		2,62	58,64	0,71	0,62	-0,48
Total	0,619	100,00				
VIFatura						
01. $\leq 20,62$		5,00	29,88	2,35	2,06	0,72

02. > 20,62 e <= 82,25		25,01	66,56	0,50	0,44	-0,82
03. > 82,25 e <= 92,19		4,99	58,55	0,71	0,62	-0,48
04. > 92,19 e <= 111,33		10,00	50,23	0,99	0,87	-0,14
05. > 111,33 e <= 132,38		10,00	44,83	1,23	1,08	0,08
06. > 132,38 e <= 171,59		15,00	39,44	1,54	1,35	0,30
07. > 171,59		30,00	34,09	1,93	1,69	0,53
Total	29,505	100,00				
VIAjuste						
01. <= 1,49		89,91	46,82	1,14	1,00	0,00
02. > 1,49 e <= 5,28		5,09	35,00	1,86	1,63	0,49
03. > 5,28		5,00	56,76	0,76	0,67	-0,40
Total	1,985	100,00				
TotalPago						
01. <= 1,90		10,00	91,27	0,10	0,08	-2,48
02. > 1,90 e <= 40,40		20,00	75,70	0,32	0,28	-1,27
03. > 40,40 e <= 51,89		5,01	68,95	0,45	0,39	-0,93
04. > 51,89 e <= 76,15		9,99	59,66	0,68	0,59	-0,52
05. > 76,15 e <= 88,74		5,00	47,37	1,11	0,97	-0,03
06. > 88,74 e <= 100,71		5,00	37,41	1,67	1,47	0,38
07. > 100,71 e <= 141,30		15,00	26,82	2,73	2,39	0,87
08. > 141,30 e <= 178,64		10,00	18,52	4,40	3,86	1,35
09. > 178,64		20,00	14,59	5,86	5,13	1,64
Total	149,468	100,00				
VIFaturaAj						
01. Igual a zero		40,31	7,57	12,21	10,71	2,37
02. > 0,00 e <= 19,45		19,74	32,72	2,06	1,80	0,59
03. > 19,45 e <= 34,65		9,99	76,79	0,30	0,26	-1,33
04. > 34,65 e <= 55,07		9,99	96,40	0,04	0,03	-3,42
05. > 55,07		19,97	99,65	0,00	0,00	-5,79
Total	487,039	100,00				
MediaPago						
00. Missing		5,37	93,64	0,07	0,06	-2,82
01. <= 15,00		5,38	86,45	0,16	0,14	-1,98
02. > 15,00 e <= 31,55		4,08	70,48	0,42	0,37	-1,00
03. > 31,55 e <= 83,19		28,39	59,21	0,69	0,60	-0,50
04. > 83,19 e <= 90,71		4,74	46,23	1,16	1,02	0,02
05. > 90,71 e <= 97,48		4,73	37,14	1,69	1,48	0,39
06. > 97,48 e <= 110,30		9,47	30,27	2,30	2,02	0,70
07. > 110,30 e <= 222,56		33,12	25,40	2,94	2,57	0,95
08. > 222,56		4,73	44,88	1,23	1,08	0,07
Total	88,658	100,00				

MaiorVIPago						
00. Missing		5,37	93,64	0,07	0,06	-2,82
01. <= 15,00		5,21	87,51	0,14	0,13	-2,08
02. > 15,00 e <= 47,21		4,25	65,29	0,53	0,47	-0,76
03. > 47,21 e <= 101,91		14,19	51,57	0,94	0,82	-0,19
04. > 101,91 e <= 123,78		9,46	39,83	1,51	1,32	0,28
05. > 123,78 e <= 242,26		42,58	34,16	1,93	1,69	0,52
06. > 242,26		18,93	46,03	1,17	1,03	0,03
Total	61,289	100,00				
QtdeLinhas						
01. <= 1		37,57	62,78	0,59	0,52	-0,65
02. > 1		62,43	37,04	1,70	1,49	0,40
Total	25,558	100,00				
VIFatAtrasoMax						
01. <= 145,08		60,00	35,70	1,80	1,58	0,46
02. > 145,08 e <= 172,02		10,00	42,94	1,33	1,16	0,15
03. > 172,02 e <= 214,84		10,00	54,44	0,84	0,73	-0,31
04. > 214,84 e <= 250,43		5,00	65,16	0,53	0,47	-0,76
05. > 250,43 e <= 313,86		5,00	71,94	0,39	0,34	-1,07
06. > 313,86 e <= 455,93		5,00	82,10	0,22	0,19	-1,65
07. > 455,93		5,00	91,83	0,09	0,08	-2,55
Total	56,444	100,00				
SeqFatura						
01. <= 9		42,60	78,72	0,27	0,24	-1,44
02. > 9 e <= 10		8,29	50,82	0,97	0,85	-0,16
03. > 10 e <= 12		17,70	32,49	2,08	1,82	0,60
04. > 12 e <= 14		10,18	20,97	3,77	3,30	1,20
05. > 14 e <= 17		10,53	8,42	10,88	9,54	2,26
06. > 17		10,70	1,77	55,55	48,69	3,89
Total	209,352	100,00				
L1feitos						
01. <= 96,55		35,00	55,74	0,79	0,70	-0,36
02. > 96,55 e <= 163,95		20,00	48,44	1,06	0,93	-0,07
03. > 163,95 e <= 236,58		15,00	42,63	1,35	1,18	0,17
04. > 236,58		30,00	37,07	1,70	1,49	0,40
Total	9,723	100,00				
L2feitos						
01. Igual a zero		78,95	47,33	1,11	0,98	-0,03
02. > 0,00 e <= 44,80		11,05	36,93	1,71	1,50	0,40
03. > 44,80 e <= 516,14		5,00	44,85	1,23	1,08	0,07
04. > 516,14		5,00	60,31	0,66	0,58	-0,55

Total	3,331	100,00				
L3feitos						
01. Igual a zero		99,96	46,73	1,14	1,00	0,00
02. >= 0,00		0,04	5,88	16,00	14,02	2,64
Total	0,170	100,00				
L4feitos						
01. Igual a zero		99,81	46,78	1,14	1,00	0,00
02. >= 0,00		0,19	12,05	7,30	6,40	1,86
Total	0,496	100,00				
L5feitos						
01. Igual a zero		99,81	46,78	1,14	1,00	0,00
02. >= 0,00		0,19	11,11	8,00	7,01	1,95
Total	0,521	100,00				
L6feitos						
01. Igual a zero		81,68	47,81	1,09	0,96	-0,04
02. >= 0,00		18,32	41,81	1,39	1,22	0,20
Total	0,875	100,00				
L7feitos						
01. <= 3,37		20,00	25,03	2,99	2,62	0,97
02. > 3,37 e <= 6,69		9,99	28,85	2,47	2,16	0,77
03. > 6,69 e <= 10,91		10,01	34,44	1,90	1,67	0,51
04. > 10,91 e <= 16,12		10,00	40,85	1,45	1,27	0,24
05. > 16,12 e <= 27,72		14,99	47,68	1,10	0,96	-0,04
06. > 27,72 e <= 33,52		5,00	53,00	0,89	0,78	-0,25
07. > 33,52 e <= 68,46		15,00	61,40	0,63	0,55	-0,60
08. > 68,46 e <= 100,67		5,00	71,01	0,41	0,36	-1,03
09. > 100,67 e <= 186,54		5,00	81,91	0,22	0,19	-1,64
10. > 186,54		5,00	92,62	0,08	0,07	-2,66
Total	72,204	100,00				
L8feitos						
01. Igual a zero		46,51	44,94	1,23	1,07	0,07
02. > 0,00 e <= 3,56		23,50	38,37	1,61	1,41	0,34
03. > 3,56 e <= 9,85		15,00	43,90	1,28	1,12	0,11
04. > 9,85 e <= 15,83		5,00	55,01	0,82	0,72	-0,33
05. > 15,83 e <= 31,89		5,00	66,17	0,51	0,45	-0,80
06. > 31,89		5,00	83,06	0,20	0,18	-1,72
Total	19,390	100,00				
L9feitos						
01. <= 8,73		29,99	53,92	0,85	0,75	-0,29
02. > 8,73 e <= 12,02		10,01	46,55	1,15	1,01	0,01
03. > 12,02 e <= 29,96		40,00	41,48	1,41	1,24	0,21

04. > 29,96 e <= 52,86		15,00	44,59	1,24	1,09	0,09
05. > 52,86		5,00	52,01	0,92	0,81	-0,21
Total	4,631	100,00				
L10feitos						
01. <= 2,78		45,00	43,39	1,30	1,14	0,13
02. > 2,78 e <= 5,99		19,99	45,49	1,20	1,05	0,05
03. > 5,99		35,01	51,67	0,94	0,82	-0,20
Total	2,240	100,00				
L11feitos						
01. <= 0,62		14,98	57,87	0,73	0,64	-0,45
02. > 0,62 e <= 1,85		5,01	53,24	0,88	0,77	-0,26
03. > 1,85 e <= 3,78		10,02	46,90	1,13	0,99	-0,01
04. > 3,78 e <= 14,37		45,00	41,67	1,40	1,23	0,20
05. > 14,37 e <= 20,27		10,00	44,16	1,26	1,11	0,10
06. > 20,27		14,99	50,08	1,00	0,87	-0,14
Total	5,608	100,00				
L12feitos						
01. Igual a zero		99,83	46,77	1,14	1,00	0,00
02. >= 0,00		0,17	8,33	11,00	9,64	2,27
Total	0,581	100,00				
L15feitos						
01. <= 0,40		64,98	44,89	1,23	1,08	0,07
02. > 0,40 e <= 5,55		20,02	41,78	1,39	1,22	0,20
03. > 5,55 e <= 9,51		5,01	47,74	1,09	0,96	-0,04
04. > 9,51 e <= 20,41		5,00	58,31	0,71	0,63	-0,47
05. > 20,41		5,00	77,53	0,29	0,25	-1,37
Total	10,722	100,00				
L16feitos						
01. Igual a zero		98,30	46,88	1,13	0,99	-0,01
02. >= 0,00		1,70	36,87	1,71	1,50	0,41
Total	0,277	100,00				
L17feitos						
01. Igual a zero		100,00	46,71	1,14	1,00	0,00
02. >= 0,00		0,00	0,00	0,00	N/A	N/A
Total	0,000	100,00				
L18feitos						
01. Igual a zero		49,36	51,51	0,94	0,83	-0,19
02. > 0,00 e <= 1,59		10,62	44,66	1,24	1,09	0,08
03. > 1,59 e <= 7,94		20,01	39,46	1,53	1,35	0,30
04. > 7,94 e <= 35,56		15,00	42,20	1,37	1,20	0,18
05. > 35,56		5,00	46,24	1,16	1,02	0,02

Total	4,134	100,00				
L19feitos						
01. <= 0,48		55,02	52,56	0,90	0,79	-0,23
02. > 0,48 e <= 26,69		34,98	40,28	1,48	1,30	0,26
03. > 26,69		10,00	37,03	1,70	1,49	0,40
Total	6,946	100,00				
L20feitos						
01. Igual a zero		47,75	53,44	0,87	0,76	-0,27
02. > 0,00 e <= 1,68		12,25	47,09	1,12	0,98	-0,02
03. > 1,68		40,00	38,56	1,59	1,40	0,33
Total	7,852	100,00				
L21feitos						
01. <= 0,32		40,02	52,21	0,92	0,80	-0,22
02. > 0,32 e <= 0,78		4,99	48,73	1,05	0,92	-0,08
03. > 0,78 e <= 40,18		44,99	42,92	1,33	1,17	0,15
04. > 40,18		10,00	40,76	1,45	1,27	0,24
Total	3,608	100,00				
L22feitos						
01. Igual a zero		99,86	46,76	1,14	1,00	0,00
02. >= 0,00		0,14	11,86	7,43	6,51	1,87
Total	0,357	100,00				
L1recebidos						
01. <= 166,70		70,00	49,72	1,01	0,89	-0,12
02. > 166,70 e <= 225,17		10,00	42,88	1,33	1,17	0,15
03. > 225,17		20,00	38,08	1,63	1,43	0,35
Total	3,717	100,00				
L2recebidos						
01. Igual a zero		99,96	46,71	1,14	1,00	0,00
02. >= 0,00		0,04	52,94	0,89	0,78	-0,25
Total	0,002	100,00				
L3recebidos						
01. Igual a zero		99,96	46,73	1,14	1,00	0,00
02. >= 0,00		0,04	6,25	15,00	13,15	2,58
Total	0,155	100,00				
L4recebidos						
01. Igual a zero		99,78	46,78	1,14	1,00	0,00
02. >= 0,00		0,22	15,96	5,27	4,62	1,53
Total	0,411	100,00				
L5recebidos						
01. Igual a zero		99,79	46,78	1,14	1,00	0,00
02. >= 0,00		0,21	13,48	6,42	5,62	1,73

Total	0,474	100,00				
L6recebidos						
01. Igual a zero		65,00	52,38	0,91	0,80	-0,23
02. >= 0,00		35,00	36,18	1,76	1,55	0,44
Total	9,818	100,00				
L7recebidos						
01. <= 5,82		30,01	39,34	1,54	1,35	0,30
02. > 5,82 e <= 10,08		15,00	42,79	1,34	1,17	0,16
03. > 10,08 e <= 15,97		14,98	45,91	1,18	1,03	0,03
04. > 15,97 e <= 30,95		20,00	49,09	1,04	0,91	-0,10
05. > 30,95 e <= 50,72		10,00	52,63	0,90	0,79	-0,24
06. > 50,72 e <= 78,35		5,00	57,91	0,73	0,64	-0,45
07. > 78,35		5,00	72,54	0,38	0,33	-1,10
Total	10,555	100,00				
L8recebidos						
01. Igual a zero		77,69	48,63	1,06	0,93	-0,08
02. > 0,00 e <= 3,30		17,31	38,36	1,61	1,41	0,34
03. > 3,30		5,00	45,77	1,18	1,04	0,04
Total	2,460	100,00				
L12recebidos						
01. Igual a zero		99,78	46,78	1,14	1,00	0,00
02. >= 0,00		0,22	13,54	6,38	5,60	1,72
Total	0,509	100,00				
L15recebidos						
01. Igual a zero		90,36	47,76	1,09	0,96	-0,04
02. >= 0,00		9,64	36,86	1,71	1,50	0,41
Total	1,712	100,00				
L16recebidos						
01. Igual a zero		99,82	46,75	1,14	1,00	0,00
02. >= 0,00		0,18	23,38	3,28	2,87	1,06
Total	0,176	100,00				
L17recebidos						
01. Igual a zero		99,97	46,72	1,14	1,00	0,00
02. >= 0,00		0,03	7,14	13,00	11,40	2,43
Total	0,125	100,00				
L18recebidos						
01. Igual a zero		42,11	52,47	0,91	0,79	-0,23
02. > 0,00 e <= 4,42		32,89	43,72	1,29	1,13	0,12
03. > 4,42		25,00	40,94	1,44	1,26	0,23
Total	4,085	100,00				
L19recebidos						

01. Igual a zero		42,69	54,09	0,85	0,74	-0,30
02. > 0,00 e <= 19,33		47,31	43,16	1,32	1,15	0,14
03. > 19,33 e <= 37,27		5,00	34,98	1,86	1,63	0,49
04. > 37,27		5,00	29,07	2,44	2,14	0,76
Total	8,556	100,00				
L20recebidos						
01. Igual a zero		16,79	47,55	1,10	0,97	-0,03
02. > 0,00 e <= 0,87		13,23	37,23	1,69	1,48	0,39
03. > 0,87 e <= 2,16		14,98	44,49	1,25	1,09	0,09
04. > 2,16 e <= 7,35		30,00	55,39	0,81	0,71	-0,35
05. > 7,35 e <= 12,98		10,00	48,25	1,07	0,94	-0,06
06. > 12,98		15,00	37,97	1,63	1,43	0,36
Total	7,680	100,00				
L21recebidos						
01. <= 0,53		20,06	57,12	0,75	0,66	-0,42
02. > 0,53 e <= 1,34		9,91	48,75	1,05	0,92	-0,08
03. > 1,34 e <= 20,02		55,02	45,36	1,20	1,06	0,05
04. > 20,02 e <= 58,59		10,00	38,06	1,63	1,43	0,36
05. > 58,59		5,00	33,09	2,02	1,77	0,57
Total	6,545	100,00				
L22recebidos						
01. Igual a zero		99,01	46,94	1,13	0,99	-0,01
02. >= 0,00		0,99	24,13	3,14	2,76	1,01
Total	0,923	100,00				
Feitos_total						
01. <= 144,94		20,00	53,57	0,87	0,76	-0,27
02. > 144,94 e <= 270,72		25,00	45,93	1,18	1,03	0,03
03. > 270,72 e <= 732,55		40,00	39,95	1,50	1,32	0,28
04. > 732,55		15,00	56,90	0,76	0,66	-0,41
Total	7,053	100,00				
Recebidos_total						
01. <= 137,37		45,00	51,49	0,94	0,83	-0,19
02. > 137,37 e <= 249,26		25,01	45,34	1,21	1,06	0,05
03. > 249,26		30,00	40,68	1,46	1,28	0,25
Total	3,512	100,00				
GROUPING - CanalVendas						
BCENT		52,90	49,78	1,01	0,88	-0,12
Dealers		35,06	45,77	1,18	1,04	0,04
Outros		12,05	35,99	1,78	1,56	0,44
Total	3,154	100,00				
GROUPING - CL_CdProdDados						

00-Missing		41,96	61,94	0,61	0,54	-0,62
01-Excelente/Muito Bom		48,82	33,44	1,99	1,75	0,56
03-Bom		5,10	42,08	1,38	1,21	0,19
06-Muito mau/Horrível/Neutro		4,12	54,62	0,83	0,73	-0,32
Total	30,976	100,00				
GROUPING - CL_CdProdVoz						
01-Excelente		5,68	23,41	3,27	2,87	1,05
02-Muito Bom		49,66	33,16	2,02	1,77	0,57
04-Neutro		5,12	55,14	0,81	0,71	-0,34
05-Mau		17,40	54,89	0,82	0,72	-0,33
06-Muito mau/Horrível/Missing		22,13	74,72	0,34	0,30	-1,22
Total	53,732	100,00				
GROUPING - CL_CEP_C						
01-Excelente/Muito Bom		16,74	33,90	1,95	1,71	0,54
03-Bom		27,87	40,69	1,46	1,28	0,25
04-Neutro		15,00	46,18	1,17	1,02	0,02
05-Mau		21,59	52,47	0,91	0,79	-0,23
06-Muito mau		12,81	58,23	0,72	0,63	-0,46
07-Horrível/Missing		5,98	66,49	0,50	0,44	-0,82
Total	14,067	100,00				
GROUPING - CL_Cidade						
01-Excelente/Muito Bom		15,07	33,88	1,95	1,71	0,54
03-Bom		26,90	40,71	1,46	1,28	0,24
04-Neutro		16,58	50,40	0,98	0,86	-0,15
05-Mau		34,77	51,99	0,92	0,81	-0,21
06-Muito mau/Horrível/Missing		6,68	63,19	0,58	0,51	-0,67
Total	10,647	100,00				
GROUPING - CL_DsProdDados						
00-Missing		41,96	61,94	0,61	0,54	-0,62
01-Excelente/Muito Bom		48,82	33,44	1,99	1,75	0,56
03-Bom		5,10	42,08	1,38	1,21	0,19
06-Muito mau/Horrível/Neutro		4,12	54,62	0,83	0,73	-0,32
Total	30,976	100,00				
GROUPING - CL_DsProdVoz						
01-Excelente		5,68	23,41	3,27	2,87	1,05
02-Muito Bom		49,66	33,16	2,02	1,77	0,57
04-Neutro		5,12	55,09	0,82	0,71	-0,34
05-Mau		17,40	54,89	0,82	0,72	-0,33
06-Muito mau/Horrível/Missing		22,14	74,73	0,34	0,30	-1,22
Total	53,745	100,00				
GROUPING - CL_EstCivil						

00-Missing		36,44	49,21	1,03	0,90	-0,10
02-Muito Bom/Bom/Neutro		51,93	44,13	1,27	1,11	0,10
05-Mau		11,63	50,43	0,98	0,86	-0,15
Total	1,187	100,00				
GROUPING - CL_GpProdDados						
00-Missing		41,96	61,94	0,61	0,54	-0,62
02-Muito Bom/Bom/Neutro		58,04	35,70	1,80	1,58	0,46
Total	27,611	100,00				
GROUPING - CL_GpProdVoz						
01-Excelente		6,67	25,02	3,00	2,63	0,97
02-Muito Bom		48,77	33,13	2,02	1,77	0,57
04-Neutro/Mau		5,33	53,03	0,89	0,78	-0,25
06-Muito mau		21,80	59,83	0,67	0,59	-0,53
07-Horrível/missing		17,43	74,66	0,34	0,30	-1,21
Total	50,960	100,00				
GROUPING - CL_NAT_OCUP						
02-Muito Bom/Missing		5,84	34,57	1,89	1,66	0,51
03-Bom		13,95	42,60	1,35	1,18	0,17
04-Neutro		39,59	46,95	1,13	0,99	-0,01
05-Mau		40,62	49,63	1,01	0,89	-0,12
Total	2,388	100,00				
GROUPING - CL_OCUPACAO						
00-Missing		45,85	49,84	1,01	0,88	-0,13
01-Excelente/Muito Bom		8,53	28,51	2,51	2,20	0,79
03-Bom		11,25	39,25	1,55	1,36	0,30
04-Neutro		25,78	46,99	1,13	0,99	-0,01
06-Muito mau/Mau/Horrível		8,59	56,99	0,75	0,66	-0,41
Total	8,132	100,00				
GROUPING - CL_SubGpProdDados						
00-Missing		41,96	61,94	0,61	0,54	-0,62
02-Muito Bom		29,31	32,84	2,04	1,79	0,58
03-Bom/Neutro		28,73	38,62	1,59	1,39	0,33
Total	28,516	100,00				
GROUPING - CL_SubGpProdVoz						
01-Excelente/Muito Bom		55,44	32,16	2,11	1,85	0,61
06-Muito mau/Neutro		22,19	59,25	0,69	0,60	-0,51
07-Horrível/Missing		22,37	70,35	0,42	0,37	-1,00
Total	46,736	100,00				
GROUPING - CL_TpProdDados						
00-Missing		41,96	61,94	0,61	0,54	-0,62
02-Muito Bom/Bom		58,04	35,70	1,80	1,58	0,46

Total	27,611	100,00				
GROUPING - CL_TpProdVoz						
07-Horrível/Missing		22,37	70,35	0,42	0,37	-1,00
03-Bom		77,63	39,90	1,51	1,32	0,28
Total	27,048	100,00				
GROUPING - CL_TpResidencia						
05-Mau/Missing		46,12	49,63	1,01	0,89	-0,12
04-Neutro		53,88	44,21	1,26	1,11	0,10
Total	1,182	100,00				
GROUPING - CL_UF						
02-Muito Bom/Bom		46,31	41,94	1,38	1,21	0,19
04-Neutro		37,17	48,64	1,06	0,93	-0,08
05-Mau/Muito mau/missing		16,51	55,76	0,79	0,70	-0,36
Total	4,125	100,00				
GROUPING - Migracao						
1) Yes		48,83	41,12	1,43	1,26	0,23
2) No		51,17	52,05	0,92	0,81	-0,21
Total	4,838	100,00				
GROUPING - WhiteList						
00-Missing		46,06	63,12	0,58	0,51	-0,67
01 - 0 a 2		15,76	12,04	7,31	6,41	1,86
02 - 3 a 5		22,06	31,00	2,23	1,95	0,67
03 - Acima de 5		16,13	55,22	0,81	0,71	-0,34
Total	72,266	100,00				

Apêndice 4 – Discretização final das variáveis contínuas

Variável	Min	Max	Maus	Bons	%Col Mau	%Col Bom	%Col Total	%Linha Mau	%Linha Bom	Odds	WOE	RR	RR - IC low	RR - IC Hight
VIatura	0	21	648	1.518	3,20%	6,57%	5,00%	29,92%	70,08%	2,34	0,72	2,05	1,88	2,25
VIatura	21	82	7.211	3.624	35,62%	15,69%	25,00%	66,55%	33,45%	0,50	-0,82	0,44	0,43	0,46
VIatura	82	92	1.269	898	6,27%	3,89%	5,00%	58,56%	41,44%	0,71	-0,48	0,62	0,57	0,67
VIatura	92	111	2.176	2.157	10,75%	9,34%	10,00%	50,22%	49,78%	0,99	-0,14	0,87	0,82	0,92
VIatura	111	132	1.944	2.391	9,60%	10,35%	10,00%	44,84%	55,16%	1,23	0,08	1,08	1,02	1,14
VIatura	132	172	2.564	3.937	12,67%	17,05%	15,00%	39,44%	60,56%	1,54	0,30	1,35	1,29	1,41
VIatura	172	3.720	4.432	8.570	21,89%	37,11%	30,00%	34,09%	65,91%	1,93	0,53	1,69	1,64	1,75
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					
VIAjuste	0	1	18.242	20.724	90,11%	89,73%	89,91%	46,82%	53,18%	1,14	0,00	1,00	0,99	1,00
VIAjuste	2	5	772	1.434	3,81%	6,21%	5,09%	35,00%	65,00%	1,86	0,49	1,63	1,49	1,77
VIAjuste	5	1.928	1.230	937	6,08%	4,06%	5,00%	56,76%	43,24%	0,76	-0,40	0,67	0,61	0,73
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					
TotalPago	0	2	3.954	378	19,53%	1,64%	10,00%	91,27%	8,73%	0,10	-2,48	0,08	0,08	0,09
TotalPago	2	40	6.563	2.107	32,42%	9,12%	20,01%	75,70%	24,30%	0,32	-1,27	0,28	0,27	0,29
TotalPago	40	52	1.495	673	7,38%	2,91%	5,00%	68,96%	31,04%	0,45	-0,93	0,39	0,36	0,43
TotalPago	52	76	2.585	1.747	12,77%	7,56%	10,00%	59,67%	40,33%	0,68	-0,52	0,59	0,56	0,63
TotalPago	76	89	1.026	1.141	5,07%	4,94%	5,00%	47,35%	52,65%	1,11	-0,03	0,97	0,90	1,06
TotalPago	89	101	810	1.357	4,00%	5,88%	5,00%	37,38%	62,62%	1,68	0,38	1,47	1,35	1,60
TotalPago	101	141	1.744	4.757	8,61%	20,60%	15,00%	26,83%	73,17%	2,73	0,87	2,39	2,27	2,52
TotalPago	141	179	802	3.532	3,96%	15,29%	10,00%	18,50%	81,50%	4,40	1,35	3,86	3,58	4,16
TotalPago	179	3.720	1.265	7.403	6,25%	32,05%	20,00%	14,59%	85,41%	5,85	1,64	5,13	4,85	5,43
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					
VIaturaAj	-97	0	1.322	16.147	6,53%	69,92%	40,31%	7,57%	92,43%	12,21	2,37	10,71	10,16	11,29

VIaturaAj	0	19	2.800	5.758	13,83%	24,93%	19,75%	32,72%	67,28%	2,06	0,59	1,80	1,73	1,88
VIaturaAj	19	35	3.324	1.004	16,42%	4,35%	9,99%	76,80%	23,20%	0,30	-1,33	0,26	0,25	0,28
VIaturaAj	35	55	4.172	156	20,61%	0,68%	9,99%	96,40%	3,60%	0,04	-3,42	0,03	0,03	0,04
VIaturaAj	55	2.083	8.626	30	42,61%	0,13%	19,97%	99,65%	0,35%	0,00	-5,79	0,00	0,00	0,00
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

MediaPago	0	15	2.016	316	9,96%	1,37%	5,38%	86,45%	13,55%	0,16	-1,98	0,14	0,12	0,15
MediaPago	15	32	1.246	522	6,15%	2,26%	4,08%	70,48%	29,52%	0,42	-1,00	0,37	0,33	0,41
MediaPago	32	83	7.284	5.019	35,98%	21,73%	28,39%	59,21%	40,79%	0,69	-0,50	0,60	0,59	0,62
MediaPago	83	91	949	1.104	4,69%	4,78%	4,74%	46,23%	53,77%	1,16	0,02	1,02	0,94	1,11
MediaPago	91	97	761	1.288	3,76%	5,58%	4,73%	37,14%	62,86%	1,69	0,39	1,48	1,36	1,62
MediaPago	97	110	1.242	2.861	6,14%	12,39%	9,47%	30,27%	69,73%	2,30	0,70	2,02	1,89	2,15
MediaPago	110	223	3.646	10.707	18,01%	46,36%	33,12%	25,40%	74,60%	2,94	0,95	2,57	2,49	2,66
MediaPago	223	3.675	920	1.130	4,54%	4,89%	4,73%	44,88%	55,12%	1,23	0,07	1,08	0,99	1,17
MediaPago			2.180	148	10,77%	0,64%	5,37%	93,64%	6,36%	0,07	-2,82	0,06	0,05	0,07
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

MaiorVIPago	0	15	1.975	282	9,76%	1,22%	5,21%	87,51%	12,49%	0,14	-2,08	0,13	0,11	0,14
MaiorVIPago	15	47	1.204	640	5,95%	2,77%	4,25%	65,29%	34,71%	0,53	-0,76	0,47	0,42	0,51
MaiorVIPago	47	102	3.172	2.979	15,67%	12,90%	14,19%	51,57%	48,43%	0,94	-0,19	0,82	0,79	0,86
MaiorVIPago	102	124	1.634	2.468	8,07%	10,69%	9,46%	39,83%	60,17%	1,51	0,28	1,32	1,25	1,41
MaiorVIPago	124	242	6.304	12.151	31,14%	52,61%	42,58%	34,16%	65,84%	1,93	0,52	1,69	1,65	1,73
MaiorVIPago	242	8.562	3.775	4.427	18,65%	19,17%	18,93%	46,03%	53,97%	1,17	0,03	1,03	0,99	1,07
MaiorVIPago			2.180	148	10,77%	0,64%	5,37%	93,64%	6,36%	0,07	-2,82	0,06	0,05	0,07
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

QtdeLinhas	0	1	10.223	6.060	50,50%	26,24%	37,57%	62,78%	37,22%	0,59	-0,65	0,52	0,51	0,53
QtdeLinhas	2	162	10.021	17.035	49,50%	73,76%	62,43%	37,04%	62,96%	1,70	0,40	1,49	1,47	1,51
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

VIFatAtrasoMax	0	145	9.284	16.720	45,86%	72,40%	60,00%	35,70%	64,30%	1,80	0,46	1,58	1,55	1,61
VIFatAtrasoMax	145	172	1.861	2.473	9,19%	10,71%	10,00%	42,94%	57,06%	1,33	0,15	1,16	1,10	1,23
VIFatAtrasoMax	172	215	2.359	1.974	11,65%	8,55%	10,00%	54,44%	45,56%	0,84	-0,31	0,73	0,69	0,78
VIFatAtrasoMax	215	250	1.412	755	6,97%	3,27%	5,00%	65,16%	34,84%	0,53	-0,76	0,47	0,43	0,51
VIFatAtrasoMax	250	314	1.559	608	7,70%	2,63%	5,00%	71,94%	28,06%	0,39	-1,07	0,34	0,31	0,37
VIFatAtrasoMax	314	456	1.779	388	8,79%	1,68%	5,00%	82,10%	17,90%	0,22	-1,65	0,19	0,17	0,21
VIFatAtrasoMax	456	8.783	1.990	177	9,83%	0,77%	5,00%	91,83%	8,17%	0,09	-2,55	0,08	0,07	0,09
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

SeqFatura	1	9	14.535	3.929	71,80%	17,01%	42,60%	78,72%	21,28%	0,27	-1,44	0,24	0,23	0,24
SeqFatura	10	10	1.825	1.766	9,02%	7,65%	8,29%	50,82%	49,18%	0,97	-0,16	0,85	0,80	0,90
SeqFatura	11	12	2.493	5.179	12,31%	22,42%	17,70%	32,49%	67,51%	2,08	0,60	1,82	1,74	1,90
SeqFatura	13	14	925	3.487	4,57%	15,10%	10,18%	20,97%	79,03%	3,77	1,20	3,30	3,08	3,54
SeqFatura	15	17	384	4.179	1,90%	18,09%	10,53%	8,42%	91,58%	10,88	2,26	9,54	8,61	10,57
SeqFatura	18	23	82	4.555	0,41%	19,72%	10,70%	1,77%	98,23%	55,55	3,89	48,69	39,17	60,53
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L1feitos	0	97	8.454	6.713	41,76%	29,07%	35,00%	55,74%	44,26%	0,79	-0,36	0,70	0,68	0,71
L1feitos	97	164	4.199	4.470	20,74%	19,35%	20,00%	48,44%	51,56%	1,06	-0,07	0,93	0,90	0,97
L1feitos	164	237	2.772	3.730	13,69%	16,15%	15,00%	42,63%	57,37%	1,35	0,17	1,18	1,13	1,23
L1feitos	237	104.501	4.819	8.182	23,80%	35,43%	30,00%	37,07%	62,93%	1,70	0,40	1,49	1,44	1,53
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L2feitos	0	0	16.197	18.021	80,01%	78,03%	78,95%	47,33%	52,67%	1,11	-0,03	0,98	0,97	0,98
L2feitos	0	45	1.768	3.019	8,73%	13,07%	11,05%	36,93%	63,07%	1,71	0,40	1,50	1,42	1,58
L2feitos	45	516	972	1.195	4,80%	5,17%	5,00%	44,85%	55,15%	1,23	0,07	1,08	0,99	1,17
L2feitos	516	235.234	1.307	860	6,46%	3,72%	5,00%	60,31%	39,69%	0,66	-0,55	0,58	0,53	0,63
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L3feitos	0	0	20.243	23.079	100,00%	99,93%	99,96%	46,73%	53,27%	1,14	0,00	1,00	1,00	1,00
L3feitos	1	125	1	16	0,00%	0,07%	0,04%	5,88%	94,12%	16,00	2,64	14,02	1,86	105,75
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L4feitos	0	0	20.234	23.022	99,95%	99,68%	99,81%	46,78%	53,22%	1,14	0,00	1,00	1,00	1,00
L4feitos	0	245	10	73	0,05%	0,32%	0,19%	12,05%	87,95%	7,30	1,86	6,40	3,31	12,39
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L5feitos	0	0	20.235	23.023	99,96%	99,69%	99,81%	46,78%	53,22%	1,14	0,00	1,00	1,00	1,00
L5feitos	0	110	9	72	0,04%	0,31%	0,19%	11,11%	88,89%	8,00	1,95	7,01	3,51	14,02
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L6feitos	0	0	16.925	18.476	83,61%	80,00%	81,68%	47,81%	52,19%	1,09	-0,04	0,96	0,95	0,97
L6feitos	0	11.090	3.319	4.619	16,39%	20,00%	18,32%	41,81%	58,19%	1,39	0,20	1,22	1,17	1,27
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L7feitos	0	3	2.170	6.498	10,72%	28,14%	20,00%	25,03%	74,97%	2,99	0,97	2,62	2,51	2,75
L7feitos	3	7	1.249	3.081	6,17%	13,34%	9,99%	28,85%	71,15%	2,47	0,77	2,16	2,03	2,30
L7feitos	7	11	1.494	2.844	7,38%	12,31%	10,01%	34,44%	65,56%	1,90	0,51	1,67	1,57	1,77
L7feitos	11	16	1.771	2.564	8,75%	11,10%	10,00%	40,85%	59,15%	1,45	0,24	1,27	1,20	1,34
L7feitos	16	28	3.098	3.399	15,30%	14,72%	14,99%	47,68%	52,32%	1,10	-0,04	0,96	0,92	1,01
L7feitos	28	34	1.149	1.019	5,68%	4,41%	5,00%	53,00%	47,00%	0,89	-0,25	0,78	0,72	0,84
L7feitos	34	68	3.993	2.510	19,72%	10,87%	15,00%	61,40%	38,60%	0,63	-0,60	0,55	0,53	0,58
L7feitos	68	101	1.538	628	7,60%	2,72%	5,00%	71,01%	28,99%	0,41	-1,03	0,36	0,33	0,39
L7feitos	101	187	1.775	392	8,77%	1,70%	5,00%	81,91%	18,09%	0,22	-1,64	0,19	0,17	0,22
L7feitos	187	21.367	2.007	160	9,91%	0,69%	5,00%	92,62%	7,38%	0,08	-2,66	0,07	0,06	0,08
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L8feitos	0	0	9.058	11.098	44,74%	48,05%	46,51%	44,94%	55,06%	1,23	0,07	1,07	1,05	1,10
L8feitos	0	4	3.908	6.277	19,30%	27,18%	23,50%	38,37%	61,63%	1,61	0,34	1,41	1,36	1,46

L8feitos	4	10	2.853	3.646	14,09%	15,79%	15,00%	43,90%	56,10%	1,28	0,11	1,12	1,07	1,17
L8feitos	10	16	1.191	974	5,88%	4,22%	5,00%	55,01%	44,99%	0,82	-0,33	0,72	0,66	0,78
L8feitos	16	32	1.434	733	7,08%	3,17%	5,00%	66,17%	33,83%	0,51	-0,80	0,45	0,41	0,49
L8feitos	32	13.599	1.800	367	8,89%	1,59%	5,00%	83,06%	16,94%	0,20	-1,72	0,18	0,16	0,20
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L9feitos	0	9	7.008	5.989	34,62%	25,93%	29,99%	53,92%	46,08%	0,85	-0,29	0,75	0,73	0,77
L9feitos	9	12	2.020	2.319	9,98%	10,04%	10,01%	46,55%	53,45%	1,15	0,01	1,01	0,95	1,06
L9feitos	12	30	7.190	10.144	35,52%	43,92%	40,00%	41,48%	58,52%	1,41	0,21	1,24	1,21	1,27
L9feitos	30	53	2.899	3.603	14,32%	15,60%	15,00%	44,59%	55,41%	1,24	0,09	1,09	1,04	1,14
L9feitos	53	1.134	1.127	1.040	5,57%	4,50%	5,00%	52,01%	47,99%	0,92	-0,21	0,81	0,75	0,88
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L10feitos	0	3	8.463	11.040	41,80%	47,80%	45,00%	43,39%	56,61%	1,30	0,13	1,14	1,12	1,17
L10feitos	3	6	3.940	4.722	19,46%	20,45%	19,99%	45,49%	54,51%	1,20	0,05	1,05	1,01	1,09
L10feitos	6	1.301	7.841	7.333	38,73%	31,75%	35,01%	51,67%	48,33%	0,94	-0,20	0,82	0,80	0,84
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L11feitos	0	1	3.757	2.735	18,56%	11,84%	14,98%	57,87%	42,13%	0,73	-0,45	0,64	0,61	0,67
L11feitos	1	2	1.157	1.016	5,72%	4,40%	5,01%	53,24%	46,76%	0,88	-0,26	0,77	0,71	0,84
L11feitos	2	4	2.037	2.306	10,06%	9,98%	10,02%	46,90%	53,10%	1,13	-0,01	0,99	0,94	1,05
L11feitos	4	14	8.126	11.376	40,14%	49,26%	45,00%	41,67%	58,33%	1,40	0,20	1,23	1,20	1,25
L11feitos	14	20	1.913	2.419	9,45%	10,47%	10,00%	44,16%	55,84%	1,26	0,10	1,11	1,05	1,17
L11feitos	20	978	3.254	3.243	16,07%	14,04%	14,99%	50,08%	49,92%	1,00	-0,14	0,87	0,84	0,91
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L12feitos	0	0	20.238	23.029	99,97%	99,71%	99,83%	46,77%	53,23%	1,14	0,00	1,00	1,00	1,00
L12feitos	0	122	6	66	0,03%	0,29%	0,17%	8,33%	91,67%	11,00	2,27	9,64	4,18	22,24
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L15feitos	0	0	12.641	15.520	62,44%	67,20%	64,98%	44,89%	55,11%	1,23	0,07	1,08	1,06	1,09
L15feitos	0	6	3.624	5.051	17,90%	21,87%	20,02%	41,78%	58,22%	1,39	0,20	1,22	1,18	1,27
L15feitos	6	10	1.036	1.134	5,12%	4,91%	5,01%	47,74%	52,26%	1,09	-0,04	0,96	0,88	1,04
L15feitos	10	20	1.263	903	6,24%	3,91%	5,00%	58,31%	41,69%	0,71	-0,47	0,63	0,58	0,68
L15feitos	20	33.268	1.680	487	8,30%	2,11%	5,00%	77,53%	22,47%	0,29	-1,37	0,25	0,23	0,28
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L16feitos	0	0	19.973	22.631	98,66%	97,99%	98,30%	46,88%	53,12%	1,13	-0,01	0,99	0,99	1,00
L16feitos	0	34.677	271	464	1,34%	2,01%	1,70%	36,87%	63,13%	1,71	0,41	1,50	1,29	1,74
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L17feitos	0	0	20.244	23.094	100,00%	100,00%	100,00%	46,71%	53,29%	1,14	0,00	1,00	1,00	1,00
L17feitos	1	1	0	1	0,00%	0,00%	0,00%	0,00%	100,00%	0,00	0,00	0,00	#NÚM!	#NÚM!
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L18feitos	0	0	11.020	10.372	54,44%	44,91%	49,36%	51,51%	48,49%	0,94	-0,19	0,83	0,81	0,84
L18feitos	0	2	2.056	2.548	10,16%	11,03%	10,62%	44,66%	55,34%	1,24	0,08	1,09	1,03	1,15
L18feitos	2	8	3.422	5.251	16,90%	22,74%	20,01%	39,46%	60,54%	1,53	0,30	1,35	1,29	1,40
L18feitos	8	36	2.744	3.759	13,55%	16,28%	15,00%	42,20%	57,80%	1,37	0,18	1,20	1,15	1,26
L18feitos	36	5.269	1.002	1.165	4,95%	5,04%	5,00%	46,24%	53,76%	1,16	0,02	1,02	0,94	1,11
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L19feitos	0	0	12.532	11.312	61,90%	48,98%	55,02%	52,56%	47,44%	0,90	-0,23	0,79	0,78	0,80
L19feitos	0	27	6.107	9.054	30,17%	39,20%	34,98%	40,28%	59,72%	1,48	0,26	1,30	1,27	1,33
L19feitos	27	2.670	1.605	2.729	7,93%	11,82%	10,00%	37,03%	62,97%	1,70	0,40	1,49	1,41	1,58
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L20feitos	0	0	11.058	9.635	54,62%	41,72%	47,75%	53,44%	46,56%	0,87	-0,27	0,76	0,75	0,78
L20feitos	0	2	2.500	2.809	12,35%	12,16%	12,25%	47,09%	52,91%	1,12	-0,02	0,98	0,94	1,04
L20feitos	2	3.078	6.686	10.651	33,03%	46,12%	40,00%	38,56%	61,44%	1,59	0,33	1,40	1,36	1,43

Total 20.244 23.095 100,00% 100,00% 100,00% 46,71% 53,29%

L21feitos	0	0	9.056	8.290	44,73%	35,90%	40,02%	52,21%	47,79%	0,92	-0,22	0,80	0,78	0,82
L21feitos	0	1	1.053	1.108	5,20%	4,80%	4,99%	48,73%	51,27%	1,05	-0,08	0,92	0,85	1,00
L21feitos	1	40	8.369	11.130	41,34%	48,19%	44,99%	42,92%	57,08%	1,33	0,15	1,17	1,14	1,19
L21feitos	40	2.490	1.766	2.567	8,72%	11,11%	10,00%	40,76%	59,24%	1,45	0,24	1,27	1,20	1,35
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L22feitos	0	0	20.237	23.043	99,97%	99,77%	99,86%	46,76%	53,24%	1,14	0,00	1,00	1,00	1,00
L22feitos	0	395	7	52	0,03%	0,23%	0,14%	11,86%	88,14%	7,43	1,87	6,51	2,96	14,33
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L1recebidos	0	167	15.085	15.253	74,52%	66,04%	70,00%	49,72%	50,28%	1,01	-0,12	0,89	0,88	0,90
L1recebidos	167	225	1.858	2.475	9,18%	10,72%	10,00%	42,88%	57,12%	1,33	0,15	1,17	1,10	1,24
L1recebidos	225	29.372	3.301	5.367	16,31%	23,24%	20,00%	38,08%	61,92%	1,63	0,35	1,43	1,37	1,48
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L2recebidos	0	0	20.235	23.087	99,96%	99,97%	99,96%	46,71%	53,29%	1,14	0,00	1,00	1,00	1,00
L2recebidos	0	526	9	8	0,04%	0,03%	0,04%	52,94%	47,06%	0,89	-0,25	0,78	0,30	2,02
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L3recebidos	0	0	20.243	23.080	100,00%	99,94%	99,96%	46,73%	53,27%	1,14	0,00	1,00	1,00	1,00
L3recebidos	0	47	1	15	0,00%	0,06%	0,04%	6,25%	93,75%	15,00	2,58	13,15	1,74	99,53
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L4recebidos	0	0	20.229	23.016	99,93%	99,66%	99,78%	46,78%	53,22%	1,14	0,00	1,00	1,00	1,00
L4recebidos	0	324	15	79	0,07%	0,34%	0,22%	15,96%	84,04%	5,27	1,53	4,62	2,66	8,02
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L5recebidos	0	0	20.232	23.018	99,94%	99,67%	99,79%	46,78%	53,22%	1,14	0,00	1,00	1,00	1,00
-------------	---	---	--------	--------	--------	--------	--------	--------	--------	------	------	------	------	------

L5recebidos	0	121	12	77	0,06%	0,33%	0,21%	13,48%	86,52%	6,42	1,73	5,62	3,06	10,33
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L6recebidos	0	0	14.756	13.414	72,89%	58,08%	65,00%	52,38%	47,62%	0,91	-0,23	0,80	0,79	0,81
L6recebidos	0	1.425	5.488	9.681	27,11%	41,92%	35,00%	36,18%	63,82%	1,76	0,44	1,55	1,50	1,59
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L7recebidos	0	6	5.117	7.890	25,28%	34,16%	30,01%	39,34%	60,66%	1,54	0,30	1,35	1,31	1,39
L7recebidos	6	10	2.782	3.720	13,74%	16,11%	15,00%	42,79%	57,21%	1,34	0,16	1,17	1,12	1,23
L7recebidos	10	16	2.980	3.511	14,72%	15,20%	14,98%	45,91%	54,09%	1,18	0,03	1,03	0,99	1,08
L7recebidos	16	31	4.256	4.413	21,02%	19,11%	20,00%	49,09%	50,91%	1,04	-0,10	0,91	0,88	0,94
L7recebidos	31	51	2.282	2.054	11,27%	8,89%	10,00%	52,63%	47,37%	0,90	-0,24	0,79	0,75	0,83
L7recebidos	51	78	1.255	912	6,20%	3,95%	5,00%	57,91%	42,09%	0,73	-0,45	0,64	0,59	0,69
L7recebidos	78	9.282	1.572	595	7,77%	2,58%	5,00%	72,54%	27,46%	0,38	-1,10	0,33	0,30	0,36
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L8recebidos	0	0	16.375	17.296	80,89%	74,89%	77,69%	48,63%	51,37%	1,06	-0,08	0,93	0,92	0,94
L8recebidos	0	3	2.878	4.625	14,22%	20,03%	17,31%	38,36%	61,64%	1,61	0,34	1,41	1,35	1,47
L8recebidos	3	4.968	991	1.174	4,90%	5,08%	5,00%	45,77%	54,23%	1,18	0,04	1,04	0,96	1,13
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L12recebidos	0	0	20.231	23.012	99,94%	99,64%	99,78%	46,78%	53,22%	1,14	0,00	1,00	1,00	1,00
L12recebidos	0	92	13	83	0,06%	0,36%	0,22%	13,54%	86,46%	6,38	1,72	5,60	3,12	10,04
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L15recebidos	0	0	18.704	20.457	92,39%	88,58%	90,36%	47,76%	52,24%	1,09	-0,04	0,96	0,95	0,96
L15recebidos	0	4.039	1.540	2.638	7,61%	11,42%	9,64%	36,86%	63,14%	1,71	0,41	1,50	1,41	1,59
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L16recebidos	0	0	20.226	23.036	99,91%	99,74%	99,82%	46,75%	53,25%	1,14	0,00	1,00	1,00	1,00
--------------	---	---	--------	--------	--------	--------	--------	--------	--------	-------------	------	------	------	------

L16recibidos	0	48	18	59	0,09%	0,26%	0,18%	23,38%	76,62%	3,28	1,06	2,87	1,70	4,87
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L17recibidos	0	0	20.243	23.082	100,00%	99,94%	99,97%	46,72%	53,28%	1,14	0,00	1,00	1,00	1,00
L17recibidos	0	98	1	13	0,00%	0,06%	0,03%	7,14%	92,86%	13,00	2,43	11,40	1,49	87,10
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L18recibidos	0	0	9.577	8.675	47,31%	37,56%	42,11%	52,47%	47,53%	0,91	-0,23	0,79	0,78	0,81
L18recibidos	0	4	6.232	8.022	30,78%	34,73%	32,89%	43,72%	56,28%	1,29	0,12	1,13	1,10	1,16
L18recibidos	4	1.058	4.435	6.398	21,91%	27,70%	25,00%	40,94%	59,06%	1,44	0,23	1,26	1,22	1,31
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L19recibidos	0	0	10.008	8.495	49,44%	36,78%	42,69%	54,09%	45,91%	0,85	-0,30	0,74	0,73	0,76
L19recibidos	0	19	8.848	11.654	43,71%	50,46%	47,31%	43,16%	56,84%	1,32	0,14	1,15	1,13	1,18
L19recibidos	19	37	758	1.409	3,74%	6,10%	5,00%	34,98%	65,02%	1,86	0,49	1,63	1,49	1,78
L19recibidos	37	1.104	630	1.537	3,11%	6,66%	5,00%	29,07%	70,93%	2,44	0,76	2,14	1,95	2,34
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L20recibidos	0	0	3.459	3.816	17,09%	16,52%	16,79%	47,55%	52,45%	1,10	-0,03	0,97	0,93	1,01
L20recibidos	0	1	2.134	3.598	10,54%	15,58%	13,23%	37,23%	62,77%	1,69	0,39	1,48	1,41	1,55
L20recibidos	1	2	2.889	3.605	14,27%	15,61%	14,98%	44,49%	55,51%	1,25	0,09	1,09	1,05	1,14
L20recibidos	2	7	7.202	5.800	35,58%	25,11%	30,00%	55,39%	44,61%	0,81	-0,35	0,71	0,69	0,73
L20recibidos	7	13	2.092	2.244	10,33%	9,72%	10,00%	48,25%	51,75%	1,07	-0,06	0,94	0,89	0,99
L20recibidos	13	1.176	2.468	4.032	12,19%	17,46%	15,00%	37,97%	62,03%	1,63	0,36	1,43	1,37	1,50
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					

L21recibidos	0	1	4.967	3.728	24,54%	16,14%	20,06%	57,12%	42,88%	0,75	-0,42	0,66	0,63	0,68
L21recibidos	1	1	2.094	2.201	10,34%	9,53%	9,91%	48,75%	51,25%	1,05	-0,08	0,92	0,87	0,98
L21recibidos	1	20	10.816	13.031	53,43%	56,42%	55,02%	45,36%	54,64%	1,20	0,05	1,06	1,04	1,07
L21recibidos	20	59	1.650	2.685	8,15%	11,63%	10,00%	38,06%	61,94%	1,63	0,36	1,43	1,35	1,51

L21recebidos	59	6.772	717	1.450	3,54%	6,28%	5,00%	33,09%	66,91%	2,02	0,57	1,77	1,62	1,93
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					
L22recebidos	0	0	20.140	22.768	99,49%	98,58%	99,01%	46,94%	53,06%	1,13	-0,01	0,99	0,99	0,99
L22recebidos	0	146	104	327	0,51%	1,42%	0,99%	24,13%	75,87%	3,14	1,01	2,76	2,21	3,43
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					
Feitos_total	0	145	4.643	4.024	22,94%	17,42%	20,00%	53,57%	46,43%	0,87	-0,27	0,76	0,73	0,79
Feitos_total	145	271	4.977	5.858	24,59%	25,36%	25,00%	45,93%	54,07%	1,18	0,03	1,03	1,00	1,07
Feitos_total	271	733	6.925	10.411	34,21%	45,08%	40,00%	39,95%	60,05%	1,50	0,28	1,32	1,29	1,35
Feitos_total	733	291.581	3.699	2.802	18,27%	12,13%	15,00%	56,90%	43,10%	0,76	-0,41	0,66	0,63	0,69
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					
Recebidos_total	0	137	10.041	9.460	49,60%	40,96%	45,00%	51,49%	48,51%	0,94	-0,19	0,83	0,81	0,84
Recebidos_total	137	249	4.914	5.923	24,27%	25,65%	25,01%	45,34%	54,66%	1,21	0,05	1,06	1,02	1,09
Recebidos_total	249	33.652	5.289	7.712	26,13%	33,39%	30,00%	40,68%	59,32%	1,46	0,25	1,28	1,24	1,32
Total			20.244	23.095	100,00%	100,00%	100,00%	46,71%	53,29%					