

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

DEPARTAMENTO DE ESTATÍSTICA

DEFINIÇÃO DE UM MODELO DE COBRANÇA (*COLLECTION SCORE*)

UTILIZANDO REGRESSÃO LOGÍSTICA MULTINOMIAL

Letícia dos Santos Lopes

Orientadora: Jandyra Maria Guimarães Fachel

Monografia apresentada para obtenção
do grau de Bacharel em Estatística

Porto Alegre, julho de 2004.

AGRADECIMENTOS

Agradeço a todos, que de alguma forma, contribuíram e acreditaram na realização deste trabalho.

À professora Jandyra Fachel pelo apoio durante o decorrer do curso e orientação deste trabalho.

Ao professor Álvaro Vigo por todos os valiosos conselhos dados sempre que precisei.

Aos demais professores do Departamento de Estatística, agradeço pelos conhecimentos que me foram passados, pois sem eles não teria sido possível chegar até aqui.

Aos meus amigos e colegas, agradeço pela amizade. Em especial à Karina, Henrique, Lauren e Marlo, que estiveram presentes em muitos momentos, tanto pessoais quanto acadêmicos.

Ao Gustavo Rossi pela oportunidade de desenvolver este estudo e ao aprendizado durante o estágio.

À minha família, principalmente, por todo apoio e compreensão, sendo essencial durante todas as dificuldades enfrentadas.

“Vá rápido quando puder.

Vá com cautela quando necessário,
mas, seja lá o que for, continue.

O importante é não parar!”

(Autor Desconhecido)

ÍNDICE

1. INTRODUÇÃO	1
1.1. Objetivos	3
2. CRÉDITO E O VAREJO	4
2.1. <i>Private Label</i>	5
3. RISCO DE CRÉDITO	7
4. INADIMPLÊNCIA E RECUPERAÇÃO DE CRÉDITOS	10
4.1. Recuperação de Créditos.....	11
5. ANÁLISE DE CRÉDITO	16
5.1. <i>Credit Score</i>	19
5.2. <i>Behaviour Score</i>	20
5.3. <i>Collection Score</i>	22
6. ANÁLISE DE REGRESSÃO.....	24
7. REGRESSÃO LOGÍSTICA.....	25
7.1. Regressão Logística Binária	26
7.2. Regressão Logística Multinomial.....	27
7.2.1. Testes de significância do modelo	28
7.2.2. Teste de significância dos coeficientes	32
7.2.3. Interpretação dos coeficientes.....	33

7.3. Validação do modelo	34
8. MODELANDO O <i>COLLECTION SCORE</i> ATRAVÉS DA REGRESSÃO	
LOGÍSTICA MULTINOMIAL	35
8.1. Definição da amostra	35
8.2. Análise exploratória dos dados	36
8.3. Variáveis estudadas.....	36
8.4. Escolha do modelo	41
8.5. Resultados.....	41
9. CONCLUSÃO.....	45
10. BIBLIOGRAFIAS.....	47
ANEXO.....	51

1. INTRODUÇÃO

O risco de crédito é a mais antiga forma de risco nos mercados financeiros. Se podemos definir crédito como a “expectativa de uma quantia em dinheiro, dentro de um espaço de tempo limitado”, então o risco de crédito é a chance de que esta expectativa não se cumpra. O risco de crédito é tão antigo quanto os empréstimos em si, o que significa que remonta a pelo menos 1800 a.C.. Essencialmente, não houve mudanças desde os tempos do Egito Antigo, agora, como naquela época, há sempre um elemento de incerteza quanto a se um tomador em particular pagará um determinado empréstimo.

Em meados da década de 80, os Estados Unidos sofreram inadimplência recorde do pagamento de empréstimos bancários e *bonds* de empresas. O mau desempenho de suas carteiras estimulou os administradores de risco de crédito a se interessarem cada vez mais por novas técnicas.

Vender sempre é um grande desafio. Em um país em recessão como o Brasil, a dificuldade fica ainda maior. Na maioria das vezes as técnicas e as facilidades oferecidas para pagamento geram um aumento nas vendas. Porém, também proporcionam o aumento da inadimplência. O que fazer? Para responder esta pergunta, sobre o que pode ser feito para reduzir a inadimplência, uma sugestão nesse trabalho é o modelo de *Collection Score* (modelo de cobrança) que analisa alguns dados da conta e prioriza os contatos com os clientes de acordo com critérios pré-estabelecidos pelo gestor de cobrança. Segundo Adilson da Silva Barros, a correta gestão da carteira de cobrança pode ser o diferencial entre a sobrevivência e a extinção da empresa. Os recursos aplicados em produção, vendas, remuneração, gastos administrativos e o esperado lucro retornam à empresa através da cobrança.

Nessa população, a de inadimplentes, é que se aplica o modelo de *Collection Score*. Este modelo fornece uma medida (score) que permite ordenar os clientes de acordo com a probabilidade de pagamento da dívida.

O fato do cliente tornar-se inadimplente não significa que a gestão estratégica do risco de crédito terminou. A inadimplência representa uma carteira com padrões específicos de comportamento que, por essa razão, merece um gerenciamento também profissional e estratégico.

A gestão da cobrança necessita ser tratada cada vez mais com profissionalismo e seriedade. A interação entre as áreas comercial, de crédito e de cobrança precisa ser total. Estas áreas devem trabalhar em conjunto, uma auxiliando a outra, para atingir o objetivo de toda a empresa, o **lucro**.

O foco deste trabalho é apresentar uma sugestão de como podem ser feitos a concepção e o desenvolvimento do *Collection Score* e mostrar uma aplicação desse modelo numa administradora de cartões *private label* do setor de varejo, utilizando como técnica a Regressão Logística Multinomial.

Atualmente, um método bastante utilizado para este tipo de análise é a regressão linear múltipla, onde todas as variáveis, dependente e independentes, são transformadas em variáveis *dummy*, dessa maneira as variáveis tornam-se categóricas.

A regressão linear é um método estatístico muito utilizado para descrever as relações ente uma variável dependente e uma ou mais variáveis independentes. Este método é usualmente empregado quando a variável dependente é contínua e é um método amplamente utilizado nos modelos de crédito. No entanto, uma suposição básica para tal método é que a variável dependente seja quantitativa, o que não ocorre nesse caso, visto que quando esta é transformada em *dummy*, passa a ser categórica. Assim, quando a variável dependente se apresenta de forma categórica, um método estatístico mais indicado é o de Regressão Logística.

É importante salientar que na Regressão Logística a variável resposta pode apresentar duas ou mais categorias. As covariáveis podem ser apresentadas em qualquer escala de medida. Na Regressão Logística tradicional a variável resposta pode assumir valores em apenas duas categorias, ou seja, a variável dependente é binária. A Regressão Logística Multinomial, que é a técnica a ser abordada neste trabalho, apresenta a variável resposta com mais de duas categorias.

1.1. Objetivos

O presente trabalho tem dois principais objetivos. Um deles é apresentar a definição do modelo de cobrança - *Collection Score* – mostrando como podem ser feitos a concepção e desenvolvimento deste modelo.

O outro é realizar um estudo com dados reais, afim de encontrar um modelo que seja capaz de identificar possíveis clientes não pagadores, utilizando como técnica estatística a Regressão Logística Multinomial.

2. CRÉDITO E O VAREJO

O Brasil é visto como possuidor de um mercado de varejo cujo potencial e dimensão são suficientemente atrativos para trazer investimentos de grandes grupos internacionais ao longo dos próximos anos. (Silveira; apud Jardim, 1996) No país, comparado a outros países, esse mercado opera com baixo grau de concentração: cerca de cem empresas controlam aproximadamente 50% dele; em países como a França e a Alemanha, esse controle é exercido por menos que dez grupos (Silveira; apud Pesciotta, 1996).

Com um volume anual de vendas superior a R\$ 100 bilhões, vendidos por cerca de um milhão de lojas, as atividades varejistas desempenham um papel de relevante importância no cenário econômico do país. Os principais setores varejistas no Brasil são: o varejo de alimentos, as revendas e lojas de carros, os postos de gasolina, as lojas de eletrodomésticos e o varejo de confecção (Parente, 2000).

Atualmente o setor varejista tem encontrado algumas dificuldades para alavancar suas vendas:

- diminuição da renda do trabalhador;
- aumento do desemprego;
- migração da renda formal para a informal;
- elevadas taxas de juros

É necessária uma queda da taxa de juros e maior alongamento dos prazos para o crédito. Dessa forma, aumenta-se a base de clientes pertencentes às classes C, D e E que possuem dificuldades de adquirir crédito, mas são clientes que consomem bastante.

Desde junho do ano passado a taxa básica de juros no Brasil vem apresentando redução. A queda já atingiu 10 pontos percentuais, pois passou de uma taxa de 26% em junho de 2003 para 16% em abril de 2004.

A partir do segundo semestre de 2003 é possível constatar que houve de fato um movimento de reativação do crédito à pessoa física. Em junho do ano passado, segundo dados do Banco Central, o saldo das operações de crédito destinadas a esses tomadores totalizava R\$ 82,3 bilhões. Em fevereiro de 2004, esse saldo já atingia R\$91,6 bilhões, ou seja, um crescimento de 11,3% no período. Da mesma

forma, a proporção desse saldo sobre o PIB evoluiu favoravelmente. Em junho do ano passado, tal taxa era de 5,3% e, em fevereiro último, ela atingiu 5,7%.

Segundo Pamplona (2004), essa melhora apenas acanhada do crédito pode ser atribuída ao fato de que as taxas cobradas dos tomadores finais ainda continuam em níveis proibitivos, muito embora tenham caído nos últimos meses. Tomando como exemplo a média das taxas prefixadas do crédito pessoal, nota-se que ela passou de 96,6% ao ano em junho de 2003 para 76,7% ao ano em fevereiro de 2004. Nesse caso, houve uma redução importante de 20 pontos percentuais, mas a nova taxa prossegue muito alta e continua impedindo que o crédito se expanda mais vigorosamente.

É preciso que a queda da taxa básica de juros da economia se intensifique e que haja uma efetiva redução dos elevados *spreads* bancários brasileiros, pois só assim as taxas de juros aos tomadores finais poderão estar em níveis aceitáveis. Dessa forma, pode ser possível presenciar uma evolução positiva marcante do volume de crédito destinado à pessoa física no Brasil

Pamplona (2004) explica que além disso, e acima de tudo, é preciso que a atividade econômica se reaqueça, que o PIB volte a crescer a mais de 3% ao ano, para que as pessoas melhorem sua renda e suas expectativas econômicas futuras, pois apenas dessa maneira elas terão estímulo para contrair dívidas e capacidade para honrá-las.

O crediário é um instrumento de grande valia para o aquecimento das vendas no varejo, mas que tem tido um enorme reflexo nas mudanças econômicas do país, pois é onde se observam maiores índices de inadimplentes configurados. O grande público que compra no varejo está sempre querendo facilidades, e um carnê para pagamento parcelado pode ser um excelente atrativo para a clientela, bastando que se tome os devidos cuidados na hora da venda (Martin, 1997).

2.1. Private Label

As grandes lojas descobriram o caminho usado pelo sistema financeiro para ganhar dinheiro no Brasil. Um estudo da consultoria Partner revela que o número de cartões de crédito emitidos por grandes empresas, como as Lojas Riachuelo, o Pão de Açúcar e a C&A, ultrapassou o das bandeiras tradicionais, como Visa e

Credicard. São 54 milhões de cartões de marcas próprias contra 46 milhões de marcas tradicionais. E as lojas estão investindo em novos negócios. Oferecem seus cartões de crédito para uso em outras lojas, vendem seguros, títulos de capitalização e até criam bancos para aproveitar melhor o negócio. “Essa é a nova fronteira do crédito no Brasil”, diz Álvaro Musa, diretor da Partner.

Pelas contas da Partner, o número de cartões de lojas cresceu 60% nos últimos três anos e já movimentava R\$ 15 bilhões por ano. Até 2006, diz Álvaro Musa, devem girar mais de R\$ 20 bilhões

Segundo Grinbaum (2003), o que sustenta essa aposta financeira de lojas de departamentos, drogarias e supermercados são três trunfos. Ao lançar marca própria, a empresa tem mais controle sobre os bancos de dados da clientela e pode fazer propaganda dirigida para seu público, sem depender do cadastro de terceiros. Também pode fazer promoções para tornar o portador do plástico um freguês fiel da loja. E há, acima de tudo, o ganho financeiro. Normalmente, uma administradora de cartões de crédito cobra de 0,5% a 1,5% sobre o faturamento. Com a marca própria, o dinheiro fica com o lojista.

3. RISCO DE CRÉDITO

Risco é uma realidade que está presente no cotidiano humano. O homem procurou sempre se defender dos riscos que estavam à sua volta, desde os tempos remotos. Recolhia-se às cavernas, desenvolvia habilidade guerreira, construía pequenas armas, utilizando, mesmo sem ter consciência, técnicas e ferramentas de proteção ao risco, de acordo com o estágio de vida e os riscos do momento.

Atualmente, a idéia de risco está associada à possibilidade de que algo ruim aconteça. De acordo com o dicionário Aurélio, a palavra risco é definida como: “perigo ou possibilidade de perigo”. A origem da palavra risco vem do italiano antigo, *risicare*, que significa “ousar”, portanto, “uma opção e não um destino”. A capacidade de administrar riscos, e com ela, a vontade de correr riscos e fazer opções ousadas são elementos-chave da energia que impulsiona o sistema econômico. Sendo o risco uma opção, então é possível medi-lo, avaliar suas conseqüências e geri-lo (Figueiredo, apud Bernstein, 1997).

A maioria das decisões de um indivíduo envolve uma escolha, é necessário se optar entre algum risco e a recompensa a ele associada. Nos negócios funciona da mesma maneira e tem uma dimensão de grande importância, pois implica no custo do capital utilizado e no retorno esperado. Administrar riscos é necessidade de qualquer empresa, por exemplo, o colapso financeiro/operacional de um banco pode causar prejuízos em cadeia e atingir boa parte da sociedade.

O risco de crédito é a mais antiga forma de risco no mercado financeiro. É conseqüência de uma transação financeira contratada entre um fornecedor de fundos (doador do crédito) e um usuário (tomador do crédito). Pode ser definido como o risco de uma contraparte, em um acordo de concessão de crédito, não honrar seu compromisso. De acordo com Ribeiro (2003), o risco de crédito diz respeito a possibilidade do cliente (de empréstimo ou operação financeira) não ser capaz de cumprir suas obrigações contratuais. Assim, a perda advinda pode ocorrer de duas maneiras distintas: uma delas é quando o cliente deixa de honrar o contrato, acarretando perda do valor do empréstimo menos o valor recuperado; essa modalidade é conhecida como risco de *default*. A outra é devido a alterações na classificação de risco do cliente, o valor de mercado do empréstimo varia; essa modalidade é denominada risco de *spread*.

Em operações financeiras, o risco pode se apresentar sob diversas formas. Os principais subtipos deste risco são (Figueiredo, 2001):

- Risco de inadimplência: risco do não-pagamento, por parte do tomador, de uma operação de crédito, empréstimo, financiamento, adiantamentos, operações de *leasing*, ou ainda a possibilidade de uma contraparte de um contrato ou emissor de um título não honrar seu crédito.
- Risco de degradação de garantia: risco de perdas em função das garantias oferecidas por um tomador deixarem de cobrir o valor de suas obrigações junto à instituição em função de desvalorização do bem no mercado, dilapidação do patrimônio empenhado pelo tomador.
- Risco de concentração de crédito: possibilidade de perdas em função da concentração de empréstimos e financiamentos em poucos setores da economia, classes de ativos, ou empréstimos elevados para um único cliente ou grupo econômico.
- Risco de degradação de crédito: perda pela queda na qualidade creditícia do tomador de crédito, emissor de um título ou contraparte de uma transação, ocasionando uma diminuição no valor de suas obrigações. Este risco pode acontecer em uma transação tipo aquisição de ações ou de títulos soberanos que podem perder valor.
- Risco soberano: risco de perdas envolvendo transações internacionais - aquisição de títulos, operações de câmbio - quando o tomador de um empréstimo ou emissor de um título não pode honrar seu compromisso por restrições do país sede.

A prática de gestão de risco foi bastante beneficiada por pesquisas na área da administração financeira. As leis da probabilidade e os conceitos de risco e retorno, aliados ao avanço da informática, proporcionaram o desenvolvimento de sistemas computacionais capazes de estimar com grande precisão o retorno de um ativo e/ou grupo de ativos. Esses sistemas profissionalizaram a área de finanças das instituições financeiras, propiciando grandes facilidades na mensuração e gerenciamento dos riscos de mercado.

Muitos banqueiros norte-americanos que sobreviveram à crise do setor têm buscado ativamente técnicas mais eficazes de gestão do risco de crédito. O Citibank, por exemplo, aplicou um sistema de inteligência artificial a um segmento de seu negócio de empréstimos comerciais. (Caouette *et al*, 1999).

Manfio (2003) afirma que tem-se pela frente um vasto mercado a ser explorado, a partir do momento em que o risco for entendido como negócio.

4. INADIMPLÊNCIA E RECUPERAÇÃO DE CRÉDITOS

No mercado financeiro, inadimplência é o termo mais utilizado para indicar a pura e simples falta de pagamento, desconsiderando suas causas ou motivos. No comércio varejista, especialmente no crediário, costuma-se denominar inadimplentes aos consumidores com pelo menos trinta dias de atraso no pagamento de suas dívidas. Contudo, podemos chamar de inadimplência qualquer dívida que tenha ultrapassado seu vencimento, pois acaba-se demonstrando o atraso no recebimento, ou a falta de pagamento.

O crescimento do sistema de crédito ao consumidor após o Plano Real, atingiu o consumidor de tal modo, que levou o país à redescoberta da inadimplência nas operações de empréstimos que, antes, era encoberta pela alta inflação e salários indexados.

Na realidade, o aumento do crédito surpreendeu o país e ao próprio consumidor, que não estavam preparados. O consumidor não havia se habituado com a estabilidade econômica e com a necessidade de conscientização do controle de suas contas. As pessoas tiveram o poder aquisitivo acrescido com a volta dos financiamentos e nem todos souberam utilizar essa vantagem dentro de suas reais condições financeiras. Como resultado, a inadimplência acabou explodindo e a maior parte das instituições não estava preparada para isto.

Percebe-se que a inadimplência no Brasil está bastante relacionada à atividade econômica. Se a economia do país está em condições satisfatórias, os índices de inadimplência são menores comparados à um ambiente econômico desfavorável. Em períodos de depressão econômica, contudo, pode se tornar uma preocupação mais séria, pois as pessoas perdem seus empregos e, conseqüentemente, atrasam seus pagamentos.

As causas da inadimplência, segundo Martin (1997), podem ser explicadas pela preocupação crescente de cada cidadão com seu bem-estar e a qualidade de vida, o que induz a querer comprar cada vez mais. Aliando-se a isso o lançamento diário de produtos novos cada vez melhores e o apelo mercadológico das campanhas publicitárias, e obtemos uma sociedade altamente consumista e por isso mesmo sujeita a dívidas de todo tipo. Para Rios (2003), a perda ou diminuição da fonte de recursos, excesso de endividamento - às vezes causado por uma oferta

exagerada de crédito, mudanças na economia, problemas com a saúde (física ou financeira) - são as causas mais comuns da inadimplência.

Para combater o aumento da inadimplência, o primeiro mecanismo de proteção adotado pelas instituições financeiras foi a elevação generalizada do *spread*, com o conseqüente aumento das taxas de juros. Esse aumento, por sua vez, contribuiu para o aumento da inadimplência, na medida em que tornou os empréstimos mais onerosos e difíceis de serem liquidados.

A ineficiência econômica – aumento da inadimplência “protegida” pelo aumento da taxa de juros, que provoca mais inadimplência – gerada nesses anos pós Plano Real, fizeram as instituições financeiras no Brasil repensarem os seus mecanismos de concessão, manutenção e recuperação de créditos. Souza (2000) explica que muitas discussões entre acadêmicos, reguladores e profissionais bancários têm demonstrado que a principal causa das perdas de crédito associadas a essa ineficiência econômica é “a falta de uma gestão mais profissional, e até mesmo estratégica, do risco de crédito, que possa melhor identificar, mensurar e administrar os vários tipos de riscos intrínsecos – risco do cliente, do tipo de operação e da concentração – à atividade de crédito”.

Nesse sentido, um grande desafio para os próximos anos no país é a gestão do risco de crédito. Assim, as instituições financeiras nacionais começam a adotar a administração estratégica do risco de crédito com a utilização de novos conceitos e instrumentos que dêem sustentação a essa nova abordagem.

4.1. Recuperação de Créditos

Apesar de tomadas as precauções, com um comportamento extremamente profissional na hora da venda, um preço competitivo e uma completa análise de crédito no cliente, muitas vezes é inevitável a falta de pagamento e conseqüente prejuízo.

Os departamentos de cobrança costumam dirigir a maior parte de suas atividades aos clientes que se utilizaram de crédito em compras a prazo ou empréstimos. A preocupação principal é o controle dos recebimentos. Considerando a atividade sob este aspecto, ainda não está se falando em inadimplentes, mas em meros clientes, pois a cobrança a que se refere também pode ser entendida como contas a receber.

Ao contrário, a expressão “recuperação de créditos” compreende especificamente as atividades relacionadas aos clientes em atraso. Sabe-se que é grande o número de empresas que mantêm departamentos próprios para esse fim, montando verdadeiras campanhas de negociação, e objetivando recuperar o maior número possível de créditos atrasados ou dados como perdidos (Martin, 1997).

Sob um outro ponto de vista, a palavra “cobrança” ainda pode ser interpretada como exigência, imposição, obrigação ao pagamento. Porém, o difícil trabalho de recuperar créditos implica em fazer muito mais do que apenas exigir.

Num mercado cada vez mais competitivo e globalizado, onde os meios de comunicação e os recursos financeiros migram de mão em mão numa velocidade surpreendente, as empresas e as pessoas se deparam com um de seus principais limites estratégicos: a capacidade de negociar.

Nessa conjuntura, muitas empresas se preocuparam tanto em vender que cometeram desmedidamente um dos principais erros em gestão de negócios, que é a falta de uma análise de crédito baseada em métodos científicos, buscando quando muito a verificação se o tomador de recursos possui garantias necessárias, quando o fundamental nessa ciência é identificar a capacidade de pagamento no futuro, pois é lá que ocorrerá ou não a solvência.

Como a própria palavra já indica, negociação envolve diálogo e entendimento, ou seja, cada parte cede um pouco para uma conclusão satisfatória para ambos. Apesar de a razão estar voltada para o credor, não é o caso de ser intransigente. Martin (1997) explica que qualquer proposta apresentada seriamente pelo devedor é, no mínimo, passível de análise e debate, pois de outra forma tratar-se-ia de uma total imposição, o que iria prejudicar, com toda a certeza, o recebimento. A essência da negociação para a inadimplência está exatamente em aproximar a capacidade de pagamento do devedor com as possibilidades de recebimento pelo credor, em todas as situações aceitáveis. Em uma pesquisa junto a algumas empresas tradicionais que possuem um volume razoável de títulos vencidos, foi possível constatar pouquíssimos resultados expressivos na recuperação dos créditos, pois os responsáveis pela cobrança faziam somente aquilo que se espera do próprio nome do departamento: cobrança. Cobrar nada mais é do que procurar o devedor e exigir o pagamento da dívida mais os juros, sem qualquer atenção aos fatos correntes no mercado financeiro. Negociar é muito mais do que isso: exige atualização,

conhecimento econômico e jurídico, ousadia e criatividade, e principalmente, exige atenção a todos os fatos e informações acerca do devedor e seu ramos de mercado.

Segundo Ettinger e Golieb (1971), a atividade em qualquer sistema de recuperação de crédito se realiza através de três etapas: negociação, protesto e cobrança judicial.

Campos Júnior (2003) diz “não existem milagres na recuperação de crédito. O que temos como aliadas são formas de relacionamento eficazes que podem se transformar em retorno persistente. Portanto, considere o devedor um potencial cliente, atenda-o, discutindo seus anseios e angústias, facilite sua vida e, principalmente, não o penalize. Não o receba no porão da sua pior filial ou trate-o como marginal. Caso insista nessa ‘estratégia do medo’, esteja certo de que, ao se recuperar, a primeira placa que ele irá avistar é a do seu concorrente.”

Tipicamente, pode-se segmentar o ciclo de cobrança em três grandes fases (Mohallem, 2003):

- Cobrança Interna: a própria empresa trata os primeiros atrasos através de um setor de cobrança interno, que funciona como uma extensão do processo de billing (faturamento). Nesta etapa, mais que recuperar a dívida, interessa recuperar o cliente de volta à normalidade (retenção). Não existe uma definição mais apurada de até quando se deve fazer a cobrança internamente, mas em média, no Brasil, os esforços internos vigoram até 30 dias de atraso.
- Cobrança Terceirizada: é transferido o serviço para empresas especializadas em recuperação de créditos. Nesse momento, constata-se um aumento dos índices de recuperação devido à especialização, qualidade de serviço e foco das empresas de cobrança. Esta etapa ocorre para as contas com mais de 30 dias de atraso e antes de serem listadas como prejuízo (normalmente com 360 dias de atraso).
- Venda de Carteira: nos casos em que todos os esforços de cobrança já foram aplicados às contas, a venda de carteira é uma opção importante de ser levada em consideração. Após todas as etapas anteriores terem se esgotado, o custo de se continuar mantendo a conta ativa em cobrança já não se justifica frente ao benefício de venda e realização imediata de receita. Hoje, no Brasil, já temos também empresas especializadas no processo de aquisição de créditos problemáticos, tanto de pessoas físicas como de

jurídicas, que já negociaram diversas carteiras. Calcula-se que o total de carteiras de créditos problemáticos já negociados no País ultrapasse R\$ 10 bilhões com promissoras expectativas de crescimento. Nos EUA, a venda de carteira já faz parte dos ciclos padronizados de cobrança e somente no mercado de cartões de créditos representa US\$ 60 bilhões por ano.

A constatação da real capacidade de recuperação de crédito e de clientes tidos como problemáticos é uma grande lição que o processo de consolidação das empresas de cobrança está oferecendo. O amadurecimento deste mercado tem proporcionado ferramentas e opções que antes não existiam e que oferecem às empresas emissoras de crédito capacidade imediata de melhorar sua rentabilidade e ter uma gestão melhor do ciclo creditício. Normalmente, os sistemas de *Credit Score* aceitam um pequeno percentual de clientes que poderão vir a ser inadimplentes. Portanto, são estes clientes que os processos de cobrança devem estar aptos a tratar com maior eficiência.

As pessoas tendem a ser boas pagadoras, e a inadimplência, na maioria dos casos, se deve a fatores alheios à vontade delas. Ações integradas de diversos departamentos, como Marketing, Crédito e Vendas, podem transformar o inadimplente em um ótimo e fiel consumidor.

Os diferentes graus de inadimplência são medidos geralmente de acordo com os dias de atraso no pagamento de uma dívida. Dentro de um prazo de até 30 dias de atraso, ou antes do vencimento de uma segunda parcela, considera-se estar numa fase amigável. Depois disso, se for o caso, pode-se até chegar a uma fase contenciosa.

Atualmente, sistemas de cobrança assistidos por computador são ferramentas adotadas nas instalações de cobrança de toda empresa internacional. De acordo com Rios (2003), consiste geralmente em superagendas, às quais todas as estratégias de cobranças são incorporadas. Estes sistemas organizam as "filas" para priorizar os acionamentos aos clientes, gerenciam o envio de cartas, ligações telefônicas, efetuam *follow-up* programado pelos supervisores e controlam a quebra de promessa de pagamento registrada para cada cliente. A organização por localizações permite uma segmentação quase perfeita da carteira de cobrança, facilitando a aplicação de estratégias diferenciadas. Sistemas assim são atualmente utilizados no Brasil e podem gerenciar simultaneamente mais de 1 milhão de contas em atraso.

Nem sempre algo que pode até parecer óbvio é conhecido e aplicado. É importante observar que não se deve gastar mais do que recupera-se. Gastar R\$ 1,05 para receber R\$ 1,00 não parece um bom negócio. É por esta simples, porém muito importante razão, que todas as estratégias e instrumentos utilizados nas atividades de cobrança têm que ser medidos passo-a-passo para se chegar à otimização dos recursos disponibilizados.

5. ANÁLISE DE CRÉDITO

A origem do crédito pode ser associada à antiga Assíria, Fenícia e Egito, onde foram descobertas diversas classes de instrumentos de crédito. Todavia, as operações de crédito, em seu verdadeiro caráter, somente foram encontradas na Grécia e em Roma (Chaia, 2003)

O termo crédito, originado do latim *creditu*, significa confiança que inspiram as boas qualidades de uma pessoa; boa fama; prazo para pagamento...” (Dicionário Brasileiro Globo), assim estabelece uma relação de confiança entre duas (ou mais) partes numa determinada operação. Dessa forma, a confiança representa uma atividade bilateral que ocorre tanto por parte do vendedor, que acredita na capacidade ou desejo do comprador de honrar os compromissos assumidos, como do adquirente em acreditar na qualidade do produto comprado.

Trata-se de algo presente no dia a dia das pessoas, particularmente, facilitando a compra e venda de serviços ou produtos. Assim, na simples compra/venda de um produto no supermercado, temos uma operação que envolve o conceito de crédito, pelos seguintes motivos: a confiança na qualidade do produto e a confiança no dinheiro (ou cheque ou cartão) utilizado na transação (Securato, 2002)

A palavra crédito pode ter mais de um significado, dependendo do contexto sob o qual esteja sendo tratada. Segundo Guimarães e Chaves Neto (2002), sob o ponto de vista meramente empresarial, a concessão de crédito significa a transferência da posse de um bem ou de uma quantia em dinheiro, mediante a promessa de pagamento futuro. De acordo com este conceito, pode-se entender o crédito à disposição de uma pessoa, física ou jurídica, como a capacidade da mesma em obter dinheiro, produtos ou serviços mediante compromisso de pagamento num determinado período de tempo.

Diariamente, há muitas decisões a serem tomadas no negócio de crédito direto ao consumidor em altos volumes. Aceitar ou rejeitar propostas de empréstimos, aumentar ou diminuir limites de crédito, autorizar ou rejeitar solicitações para compras com o cartão de crédito, etc. A utilização do controle estatístico é uma maneira prática de se gerenciar uma organização dessa natureza.

Nesse cenário, nasceram os conceitos de *Credit Score* e *Behaviour Score*. O enfoque é a utilização da estatística para prever ocorrências futuras com base na experiência passada.

Os modelos de análise de crédito são ferramentas que permitem diferenciar os grupos de uma população. Thomas (2003) explica que essa idéia de distinção entre os grupos de uma população foi introduzida por Fisher (1936), que procurou diferenciar duas variedades de íris pela mensuração do porte das plantas e distinguir a origem de crânios por meio de suas dimensões físicas. David Durand (1941) foi o primeiro a reconhecer que as mesmas técnicas poderiam ser usadas para diferenciar os bons empréstimos dos maus. Seu projeto de pesquisa foi realizado para o *National Bureau of Economic Research* dos EUA, embora não tenha sido usado para fins preditivos. Nessa época, algumas companhias financeiras e empresas de vendas por mala direta passavam por dificuldades nos seus gerenciamentos de crédito. Há muitos anos, as decisões de conceder (ou não) empréstimos, enviar (ou não) as mercadorias vinham sendo feitas por analistas de crédito. Porém, o fato desses analistas estarem sendo convocados para o serviço militar e haver uma grave falta de pessoas sem o domínio no assunto, as empresas solicitaram a esses analistas que criassem as regras de bolso (*rules-of-thumb*), que eram utilizadas para decidir a quem conceder empréstimos. Assim, essas regras passaram a ser usadas por pessoas que não tinham a experiência necessária para ajudar na tomada de decisões de crédito. Após o fim da guerra, algumas pessoas associaram esses casos e perceberam os benefícios que poderiam decorrer do uso de modelos estatísticos em decisões de crédito. A primeira consultoria nessa área foi fundada em San Francisco por Bill Fair e Earl Isaac no início da década de 50 e seus clientes, na época, eram principalmente financeiras, empresas de compras por mala direta e de varejo.

No fim dos anos 60, os bancos e demais emitentes de cartões começaram a perceber a utilidade dos modelos de crédito, através da chegada dos cartões de crédito. O número de pessoas que solicitavam cartões a cada dia exigia que a decisão de empréstimo fosse automatizada. Quando essas organizações passaram a usar o *Credit Score*, perceberam que sua capacidade preditiva era muito maior do que a de qualquer sistema julgamental e que as taxas de inadimplência caíram 50% ou mais. A aprovação dos *Equal Credit Opportunity Acts* (ECOA 1975/1976) nos EUA garantiu a aceitação total do *Credit Score*. Tais leis tornariam ilegal a

discriminação da concessão de crédito, a menos que houvesse justificativa estatística para ela. É raro os legisladores proporcionarem emprego no longo prazo para qualquer categoria, com exceção dos advogados, mas dessa vez, eles fizeram com que a análise de crédito se tornasse uma profissão em crescimento pelos 25 anos que se seguiram - e até hoje, pois o número de analistas no reino Unido dobrou nos últimos quatro anos.

Com a aceleração do desenvolvimento da informática, a partir dos anos 70, a abordagem estatística baseada na pontuação de propostas de crédito surgiu, no negócio de financiamento ao consumidor, como um dos métodos mais importantes de suporte à tomada de decisão para grandes volumes de propostas de empréstimos (Santos, 2000).

Originalmente, a abordagem para a concessão de crédito era puramente julgamental. Hoje, os modelos de crédito se baseiam em métodos estatísticos ou de pesquisa operacional. O objetivo é prever quem se tornará inadimplente e não oferecer explicações de por que a inadimplência ocorre ou responder às hipóteses sobre a relação entre inadimplência e outras variáveis econômicas e sociais.

Segundo Santos (2000), a pontuação de crédito é um instrumento estatístico desenvolvido para que o analista avalie a probabilidade de que determinado solicitante de empréstimos venha a ser um mau pagador no futuro. O sistema consiste em proceder de acordo com uma fórmula para a avaliação de cada solicitante de crédito, levando em conta um conjunto de características que a experiência de cada banco tem evidenciado como relevantes na previsão de reembolso.

Após a atribuição de valores numéricos a cada característica ou variável de risco selecionada do tomador, cada empréstimo – amparado por sistemas de pontuação – pode ser concedido ou recusado de maneira padronizada, consistente e objetiva, baseando-se nas probabilidades de reembolso calculadas.

O resultado final é o cálculo de um sistema de pontuação numérico que sumariza o perfil de crédito de cada tomador e corresponde ao risco que sua proposta significa para o banco.

De acordo com Thomas (2003), na década de 80, foram introduzidas a regressão logística e a programação linear, os dois principais baluartes dos utilizadores de *scorecards* de nossos dias. Mais recentemente, foram implementados

programas piloto com técnicas de inteligência artificial, tais como sistemas especialistas e redes neurais.

Os modelos são baseados no conceito de risco, são probabilísticos e não determinísticos, com base em fatos/casos ocorridos no passado. Assim, um modelo não serve para todos os produtos porque os fatores de risco são diferenciados. (Pereira *et al*, 2002)

5.1. Credit Score

Um modo tradicional de organizar as informações sobre a capacidade de pagamento do cliente é caracterizado pelas suas 5 dimensões, conhecidas como os 5 C's do crédito (Sousa e Chaia, 2002; Securato, 2002):

- **Caráter:** representa o histórico de pagamento dos compromissos financeiros e contratuais;
- **Capacidade:** representa o potencial financeiro para honrar os compromissos. Na pessoa física, a capacidade de pagamento está diretamente relacionada à renda;
- **Capital:** na pessoa física, o capital é o patrimônio pessoal do solicitante;
- **Colaterais:** são garantias adicionais oferecidas pelo adquirente para atendimento das demandas da empresa que concederá o crédito. Devem ser aplicados nos colaterais os mesmos procedimentos de avaliação de crédito empregados no caso do solicitante;
- **Condição:** diz respeito ao ambiente – *environment*, ou seja, a fatores macro ou microeconômicos que influenciam na concessão do crédito. São as condições econômicas vigentes e as características individuais.

O *Credit Score* é um método utilizado na análise da concessão de crédito. De acordo com Pereira *et al* (2002), é entendido como o sistema de pontuação dos dados demográficos e cadastrais de um proponente ao crédito. Com base nesse sistema, propostas são aprovadas/rejeitadas ou recomendadas/não recomendadas, dependendo do ponto de corte do modelo implementado e da função do *Credit Score* em relação à autonomia do detentor da alçada de crédito.

Caouette *et al* (1999) explica que os modelos tradicionais de *Credit Score* atribuem pesos estatisticamente predeterminados a alguns atributos dos solicitantes, para gerar um escore de crédito. Se esse escore é favorável, quando comparado a

um valor de corte, então a solicitação é aprovada. De um modo geral, utilizam-se técnicas de otimização ou técnicas estatísticas multivariadas, tais como análise discriminante ou de regressão logística, para o desenvolvimento desses modelos.

Esse modelo é usado pelos credores para determinar quão bom o risco de crédito de seu cliente é. É uma análise estatística que abrange uma série de informações do cliente, tentando formar um perfil de atuação em relação ao provento de empréstimo. Parte-se de uma verificação do histórico do cliente, passando pela sua capacidade de consumo, pela sua capacidade de liquidação de dívidas, provisão de conjunturas negativas, observação de despesas, enfim, uma série de categorias que podem ter efeito direto ou indireto na capacidade de solvência do cliente (Pereira *et al*, 2002). A idéia é trabalhar com clientes de menor risco e maior retorno. No instante em que a empresa passar a ter mais prejuízo do que ganho em virtude da não aceitação de clientes potenciais obtém-se o escore de corte para a definição de sua política de crédito.

Apesar do *Credit Score* representar um processo científico, ele não inibe a possibilidade de se recusar um bom pagador ou de se aceitar um mau pagador. Isto ocorre porque nenhum sistema de gestão de crédito consegue o total de informações relevantes na classificação do devedor e mesmo que conseguisse, o seu custo tornaria a análise economicamente inviável (Sousa e Chaia, 2003).

O *Credit Score* tem o objetivo de manter o controle da carteira, como auxiliar da decisão e não como controlador de perdas. Somente após possuir a carteira é que se verifica o quanto se quer perder. Ajuda a reduzir perdas, mas seu principal objetivo é maximizar receita (Pereira *et al*, 2002). Apenas o levantamento das informações do *Credit Score* não garante o sucesso de um modelo de gestão de crédito na empresa, devendo esta manter um contínuo acompanhamento das posições dos clientes.

5.2. Behaviour Score

Como o próprio nome define, o modelo *Behaviour Score* é um sistema de pontuação com base em análise comportamental; sua elaboração é complexa, na medida em que envolve grande quantidade de amostras de variáveis particulares

vinculadas ao comportamento dos indivíduos; por sua vez, o comportamento guarda correlação com variáveis conjunturais (Securato, 2002).

Esse modelo é formado a partir do conhecimento das operações do cliente com a instituição credora durante o seu relacionamento. As informações mais utilizadas para a determinação dos escores são: nível de utilização do limite de crédito, hábitos de pagamento, perfil de compras, escore de crédito, frequência e quantidade de inadimplência, tempo de relacionamento, etc.

Os modelos comportamentais são utilizados para aumentar a rentabilidade das contas, oferecendo uma base para renovação ou cancelamento das linhas de crédito, para autorização de compras nas lojas, para aumentar ou diminuir linhas de crédito e para a determinação de encargos diferenciados.

Novos dados de desempenho estão constantemente sendo gerados pelos mesmos clientes, isso torna a modelagem comportamental bastante dinâmica e, desta forma o modelo de *Behaviour Score* pode ser periodicamente revisado. É importante que se possa medir a eficiência de um escore de comportamento específico, comparando-o com a projeção esperada estabelecida durante a fase de desenvolvimento do modelo.

Mais do que simplesmente separar o bom crédito do mau, os modelos comportamentais permitem prever o risco cliente, ou seja, a probabilidade de que o cliente se torne inadimplente em um horizonte específico. Essas probabilidades permitem calcular as perdas esperadas para clientes com determinado perfil e dimensionar melhor a operação, definindo taxas e sugerindo diferentes níveis de garantias. Ademais, essa medida de risco é extremamente útil para a análise do risco de uma carteira de crédito (Souza, apud Douat e Sicsú, 1999).

Conhecer o risco do cliente permite a segmentação do portfólio em conjuntos de desempenhos parecidos, de tal forma que seja possível estabelecer estratégias de atuação diferentes para cada segmento, seja do ponto de vista da gestão financeira do risco – minimizar perdas de crédito ou aumentar receitas – seja do ponto de vista de marketing – aumentar o portfólio para públicos selecionados ou reter clientes de alta rentabilidade (Souza, 2000).

5.3. *Collection Score*

A identificação e a quantificação do risco de clientes tornarem-se inadimplentes têm sido os objetivos dos modelos de *Credit Score* e *Behaviour Score*, apresentados anteriormente. Dessa forma, analisa-se as probabilidades de inadimplência.

Com a finalidade de identificar a probabilidade de pagamento de clientes já inadimplentes tem-se o modelo de *Collection Score* (escore de cobrança). O objetivo é fornecer, através do escore (pontuação), uma medida que permita ordenar os clientes numa escala de menor para maior probabilidade de pagamento.

Assim, a população alvo é formada por aqueles que não efetuaram o pagamento no prazo combinado com a instituição credora. Com isso dá-se prioridade às contas mais difíceis de serem cobradas, reduzindo o número de telefonemas desnecessários (quando cobrança telefônica), o atrito com bons clientes (bons pagadores) e aumentando a relação custo x benefício da infra-estrutura utilizada para gerenciar o contas a receber.

É uma ferramenta que permite prever perdas a partir da probabilidade de pagamento de clientes já inadimplentes – ou, de forma complementar, pela probabilidade de ultrapassarem 30 dias de atraso, 60 dias, e assim por diante – e, com isso, definir recursos e estratégia de cobrança de forma mais eficiente (Souza; 2000).

Sabe-se que é mais interessante identificar o problema da inadimplência e criar mecanismos de prevenção do que atuar no efeito. Porém, a inadimplência é uma realidade em todas as operações de crédito ao consumidor. No Brasil, ainda não se nota muito interesse nesse sentido. As discussões e a literatura disponível sobre administração e tecnologia de crédito não têm demonstrado muito interesse e preocupação com esse tema.

Desde a década de 80 é comum a utilização de modelos de cobrança nos EUA, em particular em empresas de cartões de crédito, e os resultados mostram que o contato antecipado com clientes (contas) com alto risco de perda (*charge-off*) ajuda a diminuir a inadimplência e a reduzir as perdas nas contas que, de fato, tornaram-se *charge-offs*.

Os segmentos mais utilizados para o gerenciamento da inadimplência são as faixas de atraso: de 0 a 30 dias, de 61 a 90 dias e acima de 90 dias. Por reduzir a receita de financiamento da empresa, não é interessante que esta exerça alguma estratégia de cobrança sobre clientes que pagam com poucos dias de atraso. No entanto, Souza (2000) diz que é imprescindível que a empresa conheça aqueles clientes inadimplentes que têm alta probabilidade de não pagar o seu saldo devedor, para que seja possível estabelecer uma estratégia de atuação sobre eles. O objetivo, nesse caso, passa a ser o de evitar as perdas de crédito, uma vez que a não atuação sobre essas pessoas pode implicar em rolamento para faixas superiores de atraso necessitando aumentar o nível de provisão para devedores duvidosos.

Após a publicação do desenvolvimento do *Collection Score*, a Serasa passou a comercializa-lo para as instituições financeiras e de empréstimos. A administradora de cartões de crédito objeto do estudo de Souza (2000) tem feito uso destes conceitos e os resultados têm sido satisfatórios em relação às perdas de crédito.

A aplicação do modelo de *Collection Score* será apresentada com detalhes no capítulo 8.

6. ANÁLISE DE REGRESSÃO

A análise de regressão ocupa-se do estudo da dependência de uma variável, a variável dependente, em relação a uma ou mais variáveis, as variáveis independentes, com o objetivo de estimar e prever o valor médio da resposta em termos dos valores conhecidos ou fixos das variáveis explicativas.

O modelo de regressão linear múltipla é aquele que apresenta a variável resposta contínua. Este modelo necessita a aceitação de dez hipóteses, sendo que três delas podem ser consideradas como as mais importantes:

- dado o valor da variável independente, o valor médio do termo de erro aleatório u_i deve ser igual a zero;
- homocedasticidade - dado o valor da variável independente, a variância de u_i é a mesma para todas as observações;
- não existência de multicolinearidade perfeita, ou seja, não há relações lineares perfeitas entre as variáveis explicativas.

Dessa forma, quando se omitem do modelo variáveis importantes ou quando se formulam hipóteses estocásticas equivocadas sobre as variáveis do modelo, a validade da interpretação da regressão estimada é altamente questionável.

Citou-se anteriormente que modelos de crédito estimados pela técnica de regressão linear, apresentam todas as variáveis transformadas em *dummy*, ou seja, nenhuma variável contínua. Verificou-se que esta técnica exige que, pelo menos, a variável resposta seja contínua, caso contrário, os resultados são bastante duvidosos.

Contudo, este trabalho propõe um modelo estimado através da Regressão Logística, sendo esta útil na análise de dados categóricos.

7. REGRESSÃO LOGÍSTICA

Decisões são bastante frequentes na vida de um ser humano e, muitas delas, acabam envolvendo dicotomias, ou seja, decidir entre duas opções. Por exemplo, decidir entre ir à praia ou serra, direita ou esquerda. Ainda, muitas vezes se espera por uma resposta como sucesso ou fracasso em um tratamento médico, a aprovação ou não em um exame.

Assim, o modelo de Regressão Logística é útil para se expressar a relação entre uma variável dependente (resposta) categórica e uma ou mais variáveis independentes, que podem ser quantitativas ou categóricas, pois permite estimar a magnitude e a direção dos efeitos preditores. Bastante utilizado na área de modelagem de crédito, o modelo de regressão linear múltipla considera a variável dependente quantitativa, sendo que há importantes suposições que não são satisfeitas ao aplicá-lo quando a resposta é categórica.

É importante entender que o objetivo de uma análise aplicando o método de Regressão Logística é o mesmo de qualquer outra técnica de modelagem usada na estatística, que é encontrar o melhor e mais parcimonioso ajuste de modelos.

Fazendo um breve histórico da metodologia de Regressão Logística, Cox e Snell (1989) e Hosmer e Lemeshow (1989) concordam que esta técnica ganhou reconhecimento com o trabalho de Truett, Cornfield e Kennel (1967), que analisava o risco de doença coronária em um grande projeto conhecido por "*Framingham heart study*". Porém, não se sabe exatamente em qual ano foi utilizado pela primeira vez um modelo de Regressão Logística. No entanto, o trabalho ainda hoje é considerado um marco inicial dos estudos envolvendo Regressão Logística.

As primeiras aplicações do modelo logístico foram em estudos prospectivos de doenças coronárias e o processo de estimação de parâmetros sob a suposição de normalidade, que se torna desnecessária quando a estimação é feita por máxima verossimilhança via métodos numéricos (Bittencourt apud McLachlan, 1992).

Bittencourt (2002) afirma que a literatura sobre Regressão Logística é muito vasta, tendo apresentado um crescimento muito rápido. Além das inúmeras aplicações na área da saúde, a Regressão Logística também tem sido utilizada no campo da econometria, administração e educação.

7.1. Regressão Logística Binária

De acordo com Hosmer e Lemeshow (1989), muitas distribuições têm sido propostas para serem utilizadas em análises de uma variável resposta binária. Uma das principais razões para a escolha da distribuição logística é que, do ponto de vista matemático, a função é de utilização fácil e flexível.

Seja Y a variável resposta que pode assumir só dois valores, representados por sucesso ($Y = 1$) e fracasso ($Y = 0$). O valor esperado de Y é dado por

$$E(Y) = P(Y = 1) = \pi$$

que é a probabilidade de ocorrência de um evento ($Y = 1$).

A distribuição condicional da variável resposta Y segue uma binomial com probabilidade dada pela média condicional $\pi(x) = E(Y | x)$. Assim, a probabilidade de sucesso da variável dependente Y , dado o vetor das variáveis independentes $x = (x_1, x_2, \dots, x_p)$, é representado por $P(Y = 1 | x) = \pi(x)$ e, conseqüentemente, $P(Y = 0 | x) = 1 - \pi(x)$ é a probabilidade de fracasso.

A forma tradicional do modelo de Regressão Logística é

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

sendo $g(x)$ a transformação *logit*, uma função linear nos parâmetros, contínua podendo variar de $-\infty$ a $+\infty$, dada pela equação

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

onde β_i são os parâmetros do modelo, estimados pelo método da máxima verossimilhança. Para o leitor mais interessado no procedimento de estimação dos parâmetros do modelo logístico, recomenda-se Carballo (2002).

Graficamente, a relação entre uma única variável x e a função *logit* $g(x)$, apresenta um comportamento na forma de um S, sendo este um formato característico do modelo logístico.

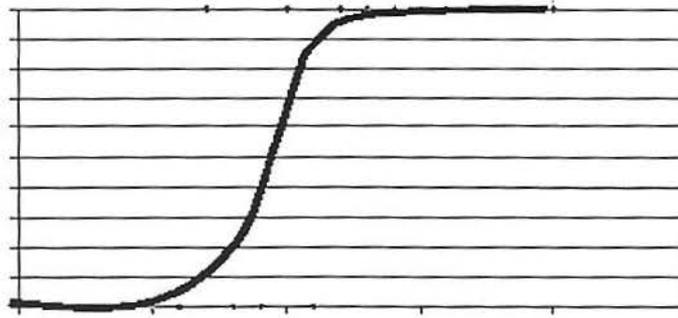


Figura 1: Função logística

A maneira mais usual de interpretar os coeficientes do modelo logístico é através da razão de chances, em inglês, *odds ratio*.

Num modelo com variável resposta e uma única covariável binárias, a chance da resposta estar presente entre indivíduos com $x=1$ é definida como $\pi(1)/[1-\pi(1)]$. Então, a razão de chances, denotada por ψ , é definida como segue:

$$\psi = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

e, conseqüentemente

$$\psi = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

O foco deste trabalho não está na Regressão Logística Binária, portanto para uma leitura detalhada dessa metodologia sugere-se, por exemplo, Hosmer e Lemeshow (1989) e Agresti (1990).

7.2. Regressão Logística Multinomial

Como já visto, o modelo de Regressão Logística é mais frequentemente aplicado para estudar a relação entre uma variável resposta dicotômica e um conjunto de covariáveis, mas com algumas modificações é possível utilizá-lo quando

a variável resposta possui mais de duas categorias, ou seja, Y é uma variável politômica.

A extensão do modelo de Regressão Logística binário para o modelo de Regressão Logística Multinomial, ou politômico, é facilitado quando a variável dependente apresenta apenas três categorias.

De início, assume-se que a variável resposta Y , com três categorias, é codificada em 0, 1 e 2. Lembrando que o modelo logístico para uma resposta binária é parametrizado em termos do *logit* de $Y=1$ em relação a $Y=0$. Assim, para uma variável resposta com três níveis tem-se duas funções *logit*: a razão entre $Y=1$ e $Y=0$ e a razão entre $Y=2$ e $Y=0$, sendo assumida a categoria $Y=0$ como referência, embora qualquer uma possa ser utilizada para essa finalidade. Geralmente a categoria de referência a ser utilizada é aquela de menor risco quando o “sucesso” na resposta representa risco ou a ausência dele.

O modelo logístico com variável resposta politômica é construído através do ajuste simultâneo de $k-1$ modelos de Regressão Logística Binária. Assim são estimados $k-1$ vetores de parâmetros $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{ip}]$, correspondentes a $k-1$ categorias da variável resposta. Dessa forma, tem-se $k-1$ comparações com a categoria de referência escolhida.

Seja $x = (x_0, x_1, x_2, \dots, x_p)$ o vetor das p covariáveis do modelo, de dimensão $p+1$, onde $x_0 = 1$. As duas funções *logit* são representadas por:

$$g_1(x) = \ln \left[\frac{P(Y = 1 | x)}{P(Y = 0 | x)} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p$$

$$g_2(x) = \ln \left[\frac{P(Y = 2 | x)}{P(Y = 0 | x)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p$$

Aqui os parâmetros também são estimados por máxima verossimilhança. Sendo que os parâmetros β_{ij} se ajustam uns para os outros. O cálculo da função de verossimilhança do modelo pode ser encontrado em Hosmer e Lemeshow (1989).

Partindo das funções *logit*, calcula-se as probabilidades condicionais de cada categoria de Y dado o vetor de covariáveis x ,

$$P(Y = 0 | x) = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)}}$$

$$P(Y = 1 | x) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}}$$

$$P(Y = 2 | x) = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}}$$

7.2.1. Testes de significância do modelo

Um dos pontos importantes no processo de modelagem de dados é verificar se o modelo estimado está bem ajustado e se é adequado para fazer inferências, caso contrário os resultados não são válidos, do ponto de vista estatístico.

Assim, pode-se iniciar fazendo a verificação do ajuste do modelo. Quando as diferenças entre os valores observados e os preditos pelo modelo é mínima, então diz-se que o modelo estatístico está bem ajustado, ou seja, é um modelo adequado para descrever a relação entre a variável resposta e as covariáveis. Daí a importância da avaliação do ajuste do modelo.

No modelo de Regressão Logística Multinomial, o ajuste do modelo também está baseado na comparação dos valores observados da variável resposta com os respectivos valores preditos pelo modelo ajustado, e esta comparação é feita para cada uma das funções *logit*, de maneira análoga ao da Regressão Logística Binária.

Hosmer e Lemeshow (1989) recomendam avaliar o ajuste do modelo logístico politômico utilizando regressões logísticas individuais, ou seja, através de $k - 1$ modelos de Regressão Logística Binária. Assim, o procedimento de avaliação do ajuste de cada modelo de Regressão Logística individual pode ser feito analogamente aos métodos de um modelo logístico binária.

A seguir serão apresentados os principais métodos existentes na literatura para verificar o ajuste do modelo logístico tradicional.

Qui-quadrado de Pearson e Deviance

Inicialmente, tem-se o valor ajustado \hat{y}_j denotado por

$$m_j \hat{\pi}_j = m_j (\exp[\hat{g}(x_j)] / \{1 + \exp[\hat{g}(x_j)]\})$$

onde $\hat{g}(x_j)$ é a estimativa do *logit*.

Os valores ajustados são calculados para cada padrão de covariáveis e dependem das correspondentes probabilidades estimadas.

A equação abaixo define a estatística qui-quadrado de Pearson,

$$X^2 = \sum_{j=1}^J \left[\frac{(y_j - m_j \hat{\pi}_j)^2}{m_j \hat{\pi}_j (1 - \hat{\pi}_j)} \right]$$

A *deviance* residual é definida por

$$d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{(m_j - y_j)}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2}$$

assim, a estatística *deviance*, baseada na equação acima, é definida por

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2$$

Sob a suposição de que o ajuste do modelo está correto em todos os aspectos, as estatísticas descritas acima, X^2 e D , têm distribuição qui-quadrado com $J - (p + 1)$ graus de liberdade, onde J é o número de combinações possíveis das categorias das covariáveis. A *deviance* representa o teste da razão de verossimilhança do modelo saturado com J parâmetros comparado ao modelo ajustado com $p + 1$ parâmetros, onde p é o número de covariáveis do modelo.

Teste de Ajustamento de Hosmer – Lemeshow

Hosmer e Lemeshow (1980) e Lemeshow e Hosmer (1982) propuseram um agrupamento baseado nos valores das probabilidades estimadas, quando $J = n$. Sendo n valores de probabilidades estimadas.

Duas estratégias foram propostas, uma baseada nos percentis das probabilidades estimadas e a outra nos valores fixos dessas mesmas probabilidades. No primeiro caso, usa-se $g = 10$ (onde g é um número fixo de grupos). Sendo o primeiro grupo contendo $n_1^i = n/10$ casos com as menores probabilidades estimadas, e o último grupo contendo $n_{10}^i = n/10$ casos com as maiores probabilidades. O segundo caso, usa $g = 10$, resultando nos pontos de corte definidos pelos valores $k/10, k = 1, 2, \dots, 9$, e os grupos contendo todos os casos com probabilidades estimadas entre os pontos de corte adjacentes.

A partir disso, criaram a estatística de avaliação do ajuste denominada estatística Hosmer-Lemeshow, \hat{C} , obtida através do cálculo da estatística qui-quadrado de Pearson de uma tabela $2 \times g$, definida por

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k^i \bar{\pi}_k)^2}{n_k^i \bar{\pi}_k (1 - \bar{\pi}_k)}$$

onde n_k^i é o número da covariável padrão no k -ésimo grupo,

$$o_k = \sum_{j=1}^{n_k^i} y_j,$$

o número respostas associadas à covariável n_k^i , e $\bar{\pi}_k = \sum m_j \hat{\pi}_j / n_k^i$ é a probabilidade média estimada.

Hosmer e Lemeshow (1980) demonstraram que, quando $J = n$ e o modelo logístico ajustado é o modelo correto, a distribuição da estatística \hat{C} , é bem

aproximada pela distribuição qui-quadrado com $g - 2$ graus de liberdade, $\chi^2(g - 2)$, onde g é o número de grupos.

A vantagem desse teste é produzir um valor simples e facilmente interpretável para o ajuste de um modelo. Porém, ele apresenta duas desvantagens, uma delas é que no processo de redução dos dados através do agrupamento, pode-se perder uma variação importante do ajuste adequado para um tamanho de amostra pequeno. A outra é o fato de ainda não ter sido implementado em nenhum programa computacional. Para maior detalhamento do teste, indica-se a leitura de Hosmer e Lemeshow (1989).

7.2.2. Teste de significância dos coeficientes

Após verificar se o modelo está bem ajustado, é necessário saber se as covariáveis presentes nele produzem um impacto significativo na variável resposta, para isso testamos se os coeficientes do modelo são significativamente diferentes de zero. Isto é bastante útil na simplificação do modelo, pois as variáveis explicativas que forem detectadas como não significantes podem ser eliminadas. E sabe-se que é preferível um modelo mais simples, com menor número de covariáveis, bem ajustado, devido à praticidade na sua interpretação. Além do que manter covariáveis pouco relevantes no modelo pode prejudicar as estimativas dos outros parâmetros.

Testar a significância de um conjunto de covariáveis consiste essencialmente em verificar se o modelo que as inclui revela melhor ajuste que o modelo que não inclui nenhuma covariável, constituído apenas do intercepto β_0 . Esta verificação é dada pela comparação entre os valores observados da variável resposta e os valores preditos para cada modelo, o que inclui e o que não inclui o conjunto das covariáveis. Se as covariáveis estão presentes no modelo, então diz-se que elas são significativas se este for melhor comparado ao modelo sem as mesmas.

Na Regressão Logística, a comparação entre os valores observados e preditos está baseada no logaritmo da função de verossimilhança. Dessa forma, o teste da razão de verossimilhança é usado para testar a hipótese nula de que o efeito das covariáveis provocado na variável resposta é nulo. Abaixo está expressa a estatística G , que é a razão de verossimilhança e é comparada com a distribuição

qui-quadrado de Pearson, sendo os graus de liberdade definidos pela diferença do número de parâmetros dos dois modelos.

$$G = -2 \ln \left[\frac{\text{verossimilhança sem as covariáveis}}{\text{verossimilhança com as covariáveis}} \right]$$

Outro teste, equivalente ao citado anteriormente, é o denominado teste de Wald. O teste de Wald é obtido através da comparação da estimativa de máxima verossimilhança do coeficiente β_{ij} com o seu respectivo erro padrão. A estatística W de Wald é dada por

$$W = \frac{\hat{\beta}_{ij}}{SE(\hat{\beta}_{ij})},$$

onde $i = 1, 2, \dots, k-1$ e $j = 0, 1, \dots, p$

Sob a hipótese de que $\beta_{ij} = 0$, essa estatística segue uma distribuição normal padronizada. Vale lembrar que este é um teste usado para a verificação isolada de cada covariável.

7.2.3. Interpretação dos coeficientes

A interpretação dos parâmetros neste caso é similar ao do modelo logístico com resposta binária. No modelo de resposta dicotômica, com apenas uma covariável dicotômica, o logaritmo da razão de chances obtida através de uma tabela de contingência 2×2 equivale à estimativa do coeficiente β_i . Quando a resposta é politômica, o processo é análogo. Porém, a razão de chances é obtida pelos pares de tabelas 2×2 da variável resposta com a covariável, usando a categoria de referência.

Assumindo $Y = 0$ como a categoria de referência, o *odds ratio* da resposta j em relação a 0 para a covariável dicotômica $x = a$ em relação a $x = b$ é

$$\psi_j(a,b) = \frac{P(Y = j | x = a) / P(Y = 0 | x = a)}{P(Y = j | x = b) / P(Y = 0 | x = b)}$$

7.3. Validação do modelo

As amostras de validação são utilizadas para verificar se o modelo estimado mantém seu poder de discriminação para amostras provindas da mesma população da amostra de desenvolvimento. Se o poder de discriminação variar muito de uma amostra para outra, pode significar que o modelo não é estável ou pode estar havendo uma superestimação – *overfitting* (Rosa, 2000).

Em algumas situações é possível excluir uma subamostra dos dados e desenvolver o modelo baseado nas informações restantes. A validação é feita através da subamostra excluída, assim, o modelo encontrado é aplicado à esses dados. Outra maneira de validar os resultados, é ajustando o modelo em uma nova amostra.

Nos últimos anos, diversos melhoramentos deste procedimento têm sido implementados, mediante o emprego de técnicas de re-amostragem, tais como *bootstrapping* e *jackknifing*.

8. MODELANDO O *COLLECTION SCORE* ATRAVÉS DA REGRESSÃO LOGÍSTICA MULTINOMIAL

Uma administradora de cartões de crédito *private label* necessita de uma ferramenta que determine quais clientes necessitam de uma cobrança mais efetiva no pagamento de suas dívidas, através da identificação dos indicadores selecionados (covariáveis). Deste modo, o objetivo é criar um modelo de cobrança que associe um score de risco aos clientes com atraso superior a 10 dias.

8.1. *Definição da amostra*

Todo modelo estatístico é iniciado a partir da coleta de uma amostra que represente a população de interesse. A política da empresa estabelece que atraso superior a 10 dias deve ser encaminhado ao departamento de cobrança. Assim, a população-alvo foi constituída por clientes que tiveram atraso maior que 10 dias no pagamento de, pelo menos, uma prestação, ou seja, são os clientes que foram enviados à cobrança.

O período analisado foi o mês de Abril do ano de 2002. A amostra de desenvolvimento constitui-se de 28.799 clientes, com atrasos entre 10 e 180 dias. Atrasos superiores a 180 dias foram excluídos do estudo, pois a esses é dado um tratamento diferenciado.

A amostra de validação apresentou 12.167 clientes analisados pelo modelo ajustado.

A separação de amostra de desenvolvimento e validação foi feita através de rotina computacional, gerando-se uma variável aleatória uniformemente distribuída utilizada para alocar, ao acaso, os casos às respectivas amostras.

8.2. *Análise exploratória dos dados*

Após a definição e coleta da amostra é importante analisar de forma exploratória todas as variáveis. Verifica-se as distribuições de frequência de cada covariável e da variável resposta, buscando analisar possíveis inconsistências, presença de valores extremos (*outliers*) e dados faltantes (*missings*).

É também importante estudar a relação de cada covariável com a variável resposta, bem como a relação entre as covariáveis, evitando a multicolinearidade entre elas. Essa análise auxilia na definição de quais variáveis explicativas podem ser importantes para o modelo.

8.3. *Variáveis estudadas*

O *Collection Score* ainda é um modelo pouco estudado na área de crédito e cobrança, por essa razão nesse trabalho optou-se por uma análise simples. Assim, as variáveis estudadas são, na sua grande maioria, sócio-demográficas.

Quando se constrói modelos para aplicação na área de crédito, deve-se primar pela simplicidade do uso e da implementação da técnica, tendo em vista que nem sempre os profissionais que manusearão os modelos têm formação matemática-estatística. Desta forma, se as variáveis do modelo puderem ser mostradas em categorias, torna-se mais simples a implementação dos modelos e interpretação dos pesos relativos às categorias das variáveis.

Contudo, abaixo estão as variáveis definidas para a análise:

- **Sexo:** variável binária, dividida em feminino e masculino;
- **Estado civil:** solteiro, casado, divorciado e viúvo;
- **Escolaridade:** verificou-se que pessoas analfabetas, com 1º ou com 2º graus apresentavam o mesmo comportamento, então essa variável foi recodificada em

duas categorias. Sendo a outra categoria composta pelas pessoas com nível superior;

- **Idade:** a variável foi categorizada em três níveis: até 25 anos, de 26 a 40 anos e mais de 40 anos;
- **Código de Endereçamento Postal (CEP) residencial:** esta informação é composta de oito dígitos, porém ela foi truncada e passou a ser analisada com os três primeiros, afim de facilitar o agrupamento.
- **Código de Endereçamento Postal (CEP) profissional:** esta variável foi analisada da mesma maneira que CEP residencial;
- **Telefone residencial:** variável binária , informa se o cliente possui ou não telefone na residência;
- **Telefone residencial é celular:** se foi declarado telefone residencial, verifica-se se é um número fixo ou celular;
- **Telefone comercial:** variável binária , informa se o cliente possui ou não telefone no local de trabalho;
- **Telefone comercial é celular:** se foi declarado telefone comercial, verifica-se se é um número fixo ou celular;
- **Cargo:** as categorias dessa variável foram estabelecidas da mesma forma que CEP residencial;
- **Tempo de relacionamento:** variável categorizada em três níveis, descreve há quanto tempo o indivíduo é cliente da empresa;
- **Tempo no atual emprego:** composta por três categorias, até 2 anos, de 3 a 7 anos e mais de 7 anos;

- **Renda:** até R\$450, de R\$451 a R\$550, mais de R\$551;
- **Escore:** definiu-se duas categorias baseadas nas classes de score dos modelos de *credit score* e *behaviour score*, utilizados pela empresa;
- **Atraso em março/2002:** para esta variável verificou-se se alguma prestação havia sido paga com atraso ou ainda estava sem pagamento. Constatou-se que um pagamento com atraso de até 30 dias apresentava comportamento semelhante à sem atraso. Vale lembrar que março é o mês que antecede o mês em estudo;
- **Serviço de Proteção ao Crédito (SPC) em março/2002:** variável binária que informa se o cliente havia sido registrado ou não no SPC no mês anterior.

A variável resposta, tempo de atraso, foi segmentada em três categorias: de 11 a 30 dias, de 31 a 60 dias e de 61 a 180 dias. Sendo que quando o cliente possuía mais de uma prestação em atraso, era considerada aquela que tinha o maior tempo. A categoria de referência considerada é o atraso de 10 a 30 dias.

As variáveis a serem categorizadas baseiam-se no método da proporção entre clientes "bons" e "maus". Clientes "bons" são aqueles que se encontram na categoria de menor atraso, isto é, de 10 a 30 dias, considerados de baixo risco. Aqueles que tiveram atraso de 61 a 180 dias são ditos "maus", pois apresentam alto risco de não pagamento. Por consequência, o atraso de 31 a 60 dias é considerado um risco médio. Com isso, agrupam-se clientes com o mesmo comportamento de risco. Além desse método, foi utilizada a análise de resíduos ajustados para as tabelas de contingência de cada covariável com a variável resposta, como um auxílio na categorização, identificando assim, qual classe da resposta estava mais associada com a categoria da covariável. Para variáveis nominais, o percentual considerado como representativo da amostra foi definido como acima de 1% (valores inferiores a 1% não tiveram seu risco analisado).

Com exceção das variáveis sexo, estado civil, SPC em março/2002 e as referentes à informação de telefone, as demais foram categorizadas da forma descrita no parágrafo anterior. Como exemplo é apresentada no Anexo a tabela de contingência para a variável Tempo de Relacionamento.

Quadro 1. Descrição das categorias das covariáveis

Sexo	Feminino Masculino
Estado Civil	Solteiro Casado Divorciado Viúvo
Escolaridade	Analfabeto, 1º grau, 2º grau Nível superior
Idade	até 25 anos 26 a 40 anos acima de 40 anos
CEP residencial	Alto risco Médio risco Baixo risco
CEP profissional	Alto risco Médio risco Baixo risco
Telefone residencial	Sim Não
Telefone residencial é celular	Sim Não
Telefone comercial	Sim Não
Telefone comercial é celular	Sim Não
Cargo	Alto risco Médio risco Baixo risco
Tempo de relacionamento	até 1 ano 2 a 4 anos mais de 5 anos
Tempo no atual emprego	até 2 anos 3 a 7 anos acima de 7 anos
Renda	até R\$450 R\$451 a R\$550 acima de R\$550
Score	Alto risco Baixo risco
Atraso em março/2002	Sim Não
SPC em março/2002	Sim Não

Tabela 1. Distribuição de frequências das covariáveis, por faixa de atraso

	Atraso			p-value
	11 a 30 dias	31 a 60 dias	61 a 180 dias	
Sexo, n(%)				0,011
Masculino	5.456 (32,4)	2.899 (32,3)	1.044 (35,1)	
Feminino	11.399 (67,6)	6.069 (67,7)	1.932 (64,9)	
Estado Civil, n(%)				0,011
Solteiro	8.192 (48,6)	4.164 (46,4)	1.462 (49,2)	
Casado	6.677 (39,6)	3.709 (41,4)	1.139 (38,3)	
Divorciado	1.089 (6,5)	586 (6,5)	197 (6,6)	
Viúvo	890 (5,3)	507 (5,7)	176 (5,9)	
Escolaridade, n(%)				<0,001
1º e 2º graus/Analfabeto	15.221 (92,5)	7.978 (92,2)	2.682 (94,9)	
Superior	1.230 (7,5)	679 (7,8)	145 (5,1)	
Idade, n(%)				<0,001
até 25 anos	3.447 (20,5)	1.739 (19,4)	720 (24,2)	
26 a 40 anos	6.166 (36,6)	3.148 (35,1)	992 (33,3)	
acima de 40 anos	7.242 (43,0)	4.081 (45,5)	1.264 (42,5)	
CEP Residencial, n(%)				<0,001
Alto risco	3.508 (20,8)	2.286 (25,5)	772 (26,0)	
Médio risco	6.928 (41,1)	3.499 (39,0)	1.224 (41,2)	
Baixo risco	6.415 (38,1)	3.178 (35,5)	978 (32,9)	
CEP Profissional, n(%)				<0,001
Alto risco	1.876 (13,5)	1.267 (17,6)	431 (18,0)	
Médio risco	9.026 (65,0)	4.566 (63,3)	1.536 (64,0)	
Baixo risco	2.984 (21,5)	1.385 (19,2)	432 (18,0)	
Telefone Residencial, n(%)				<0,001
Não	1.246 (7,4)	772 (8,7)	301 (10,2)	
Sim	15.481 (92,6)	8.136 (91,3)	2.644 (89,8)	
Telefone residencial é celular, n(%)				<0,001
Sim/Não declarou	3.970 (23,7)	2.125 (23,9)	861 (29,2)	
Não	12.757 (76,3)	6.783 (76,1)	2.084 (70,8)	
Telefone Comercial, n(%)				0,001
Não	4.028 (24,2)	2.312 (26,0)	779 (26,4)	
Sim	12.650 (75,8)	6.566 (74,0)	2.171 (73,6)	
Telefone comercial é celular, n(%)				<0,001
Sim/Não declarou	4.259 (25,5)	2.462 (27,7)	822 (27,9)	
Não	12.419 (74,5)	6.416 (72,3)	2.128 (72,1)	
Cargo, n(%)				<0,001
Alto risco	4.987 (30,5)	2.796 (32,6)	995 (35,8)	
Médio risco	5.440 (33,3)	2.932 (34,2)	940 (33,8)	
Baixo risco	5.904 (36,2)	2.839 (33,1)	843 (30,3)	
Tempo no atual emprego, n(%)				<0,001
até 2 anos	3.974 (28,6)	1.890 (25,8)	736 (30,0)	
3 a 7 anos	7.070 (50,9)	3.907 (53,4)	1.304 (53,1)	
acima de 7 anos	2.833 (20,4)	1.525 (20,8)	414 (16,9)	
Renda, n(%)				<0,001
até R\$451	5.170 (30,7)	2.847 (31,8)	1.157 (38,9)	
R\$451 a R\$551	2.195 (13,0)	1.140 (12,7)	376 (12,6)	
acima de R\$551	9.484 (56,3)	4.971 (55,5)	1.440 (48,4)	
Escore - CS/BHS, n(%)				<0,001
Alto risco	8.441 (50,1)	4.508 (50,3)	2.877 (96,7)	
Baixo risco	8.414 (49,9)	4.460 (49,7)	99 (3,3)	
Atraso em março/2002, n(%)				<0,001
Superior a 31 dias	2.776 (16,5)	3.448 (38,4)	1.921 (64,5)	
Sem atraso/inferior a 30 dias	14.079 (83,5)	5.520 (61,6)	1.055 (35,5)	
SPC em março/2002, n(%)				<0,001
Sim	1.096 (6,5)	1.882 (21,0)	1.632 (54,8)	
Não	15.759 (93,5)	7.086 (79,0)	1.344 (45,2)	
Total	16.855 (58,5)	8.968 (31,1)	2.976 (10,3)	

8.4. Escolha do modelo

A decisão das variáveis a serem incluídas no modelo baseou-se em dois métodos. Um deles teve a finalidade de identificar uma possível multicolinearidade entre as covariáveis. Utilizando-se o teste de associação qui-quadrado de Pearson, verificou-se a relação entre as covariáveis. Identificando tal associação, algumas dessas variáveis tornaram-se candidatas à exclusão da análise. O outro método adotado é o proposto por Hosmer e Lemeshow (1989), onde são produzidos modelos de Regressão Logística Multinomial para cada covariável individualmente. Então, aquela covariável que não se mostrou significativa ($p - value < 0,25$) pode ser indicada a não fazer parte do modelo final. É importante ressaltar que $p - value < 0,25$ é utilizado apenas na etapa de seleção de variáveis "candidatas" ao modelo final. Quando se tem os resultados ajustados com todas as covariáveis definidas importantes, então o nível de significância a ser considerado é o comumente utilizado, $\alpha = 0,05$.

Com isso, algumas variáveis explanatórias foram sendo excluídas e a cada exclusão um novo modelo era ajustado. A comparação entre eles foi feita através dos testes de ajustes. Dessa forma, chegou-se no modelo final, avaliado como sendo o mais adequado.

Cabe ressaltar que os dados foram analisados usando o programa SPSS® versão 11.5.

8.5. Resultados

Após todas as etapas importantes terem sido avaliadas, o modelo escolhido foi aquele que inclui as seguintes covariáveis: escolaridade, idade, CEP profissional, telefone residencial, cargo, renda, atraso em março/2002, SPC em março/2002. Os valores encontrados para o ajuste do modelo que inclui essas variáveis independentes são os seguintes:

Qui-quadrado de Pearson - $X^2 = 1227,68$, com 1236 gl ($p - value = 0,561$),
Deviance - D = 1280,17, com 1236 gl ($p - value = 0,186$)

Nota-se que as duas estatísticas de ajuste aceitam a hipótese de que o modelo está bem ajustado, ao nível de significância 5%.

Todas as covariáveis foram significativas em pelo menos uma das categorias da resposta.

O principal interesse na modelagem dos dados é a criação de um índice de risco. Através dele, é possível identificar se o cliente tem alta ou baixa probabilidade de pagamento da dívida.

O índice é uma pontuação (score) calculada com base nos valores dos coeficientes estimados em cada categoria da covariável. Esses valores foram multiplicados por dez, seguindo a sugestão de Arozullah *et. al* apud Le Gall *et al* (1993). Os autores não apresentam nenhuma justificativa para tal procedimento, no entanto, imagina-se que um motivo seja a facilidade no manuseio de cálculos matemáticos.

Sendo a categoria de atraso 61 a 180 dias a que oferece maior risco, optou-se em utilizar apenas os coeficientes estimados para esse nível.

A tabela 2 apresenta os resultados encontrados.

Tabela 2. Resultados da Regressão Logística Multinomial: estimativas dos coeficientes, teste de Wald, *odds ratio*

Estimativas dos parâmetros	Beta	E.P	Wald	gl	Sig.	Exp(Beta)	IC 95% para Exp(Beta)	
							Limite inferior	Limite superior
61 - 180 dias	Intercept	-3,32	0,13	676,40	1	0,00		
	CEPPRO= AR	0,49	0,08	35,27	1	0,00	1,64	1,39 1,93
	CEPPRO=MR	0,17	0,07	7,02	1	0,01	1,19	1,05 1,35
	CEPPRO=BR							
	DTFONE=N	0,20	0,09	5,09	1	0,02	1,23	1,03 1,46
	DTFONE=S							
	CARGO=AR	0,22	0,06	11,61	1	0,00	1,24	1,10 1,41
	CARGO=MR	0,15	0,06	6,33	1	0,01	1,16	1,03 1,30
	CARGO=BR							
	IDADE=1	0,00	0,07	0,00	1	0,96	1,00	0,87 1,14
	IDADE=2	-0,08	0,06	1,73	1	0,19	0,92	0,82 1,04
	IDADE=3							
	ESCOLA=12A	0,27	0,11	6,05	1	0,01	1,31	1,06 1,62
	ESCOLA=S							
	RENDA=1	0,32	0,06	31,77	1	0,00	1,38	1,24 1,55
RENDA=2	0,06	0,08	0,51	1	0,47	1,06	0,91 1,23	
RENDA=3								
ATR1N=S	0,87	0,08	112,32	1	0,00	2,39	2,03 2,81	
ATR1N=N								
SPC1N=S	2,10	0,09	600,56	1	0,00	8,20	6,93 9,70	
SPC1N=N								
31 - 60 dias	Intercept	-1,04	0,07	241,84	1	0,00		
	CEPPRO= AR	0,39	0,05	58,61	1	0,00	1,48	1,34 1,64
	CEPPRO=MR	0,07	0,04	3,27	1	0,07	1,07	0,99 1,16
	CEPPRO=BR							
	DTFONE=N	0,05	0,06	0,73	1	0,39	1,05	0,94 1,19
	DTFONE=S							
	CARGO=AR	0,13	0,04	11,17	1	0,00	1,14	1,06 1,24
	CARGO=MR	0,15	0,04	17,71	1	0,00	1,16	1,08 1,25
	CARGO=BR							
	IDADE=1	-0,13	0,04	9,35	1	0,00	0,88	0,81 0,95
	IDADE=2	-0,07	0,04	4,30	1	0,04	0,93	0,87 1,00
	IDADE=3							
	ESCOLA=12A	-0,11	0,06	3,73	1	0,05	0,90	0,80 1,00
	ESCOLA=S							
	RENDA=1	0,07	0,04	3,91	1	0,05	1,07	1,00 1,15
RENDA=2	-0,01	0,05	0,02	1	0,89	0,99	0,90 1,09	
RENDA=3								
ATR1N=S	0,91	0,04	430,36	1	0,00	2,49	2,28 2,71	
ATR1N=N								
SPC1N=S	0,57	0,06	95,21	1	0,00	1,77	1,58 1,99	
SPC1N=N								

Variáveis *dummy* foram criadas para que todas as categorias das covariáveis fossem pontuadas.

Quando um modelo de crédito é analisado com uma variável resposta binária, um teste bastante utilizado para a verificação da separação entre os grupos é o teste de Kolmogorov-Smirnov (KS). Segundo Souza apud Cordell *et al* (1998), modelos dessa natureza devem ter separação entre 40% e 60% para serem considerados viáveis. É bastante comum encontrar bons modelos com separação de 40% a 50%.

Como a variável resposta do modelo apresenta três categorias, optou-se utilizar a técnica de Análise Discriminante para analisar a separação dos grupos de atraso.

Para esta técnica parte-se de grupos definidos *a priori* e verifica se o vetor de variáveis analisadas consegue discriminar tais grupos. A análise se dá através da estimação de $k - 1$ funções discriminantes, onde k é o número de grupos. Maiores detalhes a respeito da Análise Discriminante, ver Braga (2003).

Assim, a variável analisada pela Discriminante foi o escore criado na etapa anterior pela análise de Regressão Logística Multinomial, partindo dos grupos de atraso 11 a 30 dias, 31 a 60 dias e 61 a 180 dias. Foi possível identificar a boa separação entre os três grupos, uma vez que o percentual de classificação correta foi igual a 60.4%. Na amostra de validação do modelo, a discriminação entre os grupos foi igual a 60.2% .

9. CONCLUSÃO

Este trabalho mostrou a importância do modelo de *Collection Score* na administração do risco de crédito, pois sabe-se que a inadimplência é uma realidade em qualquer empresa que conceda crédito.

O gerenciamento da carteira de inadimplentes é uma necessidade real, dessa forma o trabalho buscou demonstrar o potencial do modelo de cobrança como ferramenta para a administração dessa carteira, dentro de uma abordagem profissional e estratégica.

Embora esse potencial ainda seja pouco explorado, a expectativa é de que haja um avanço nesse sentido no Brasil, em virtude da própria necessidade aqui discutida e das mudanças rápidas que vêm ocorrendo nas tecnologias de informação, bem como no desenvolvimento de técnicas estatísticas para exploração e análise de dados (*data mining*).

Contudo, o interesse principal era modelar os dados de clientes inadimplentes utilizando a técnica estatística de Regressão Logística Multinomial, estabelecendo um modelo de *Collection Score*.

A modelagem de dados via Regressão Logística tem sido bastante difundida na área biomédica, sobretudo em estudos epidemiológicos. No entanto, a área de crédito parece ter encontrado nessa técnica uma ferramenta bastante importante quando se quer medir algum tipo de risco. Dessa forma, a análise dos dados através da Regressão Logística é bastante importante visto que ela é a mais indicada no caso de variáveis categóricas, enquanto que a Análise de Regressão Múltipla tem seus resultados questionáveis quando aplicada nesse tipo de dados. Assim, o trabalho espera contribuir no avanço da modelagem de crédito, mostrando a importância de se analisar dados fazendo uso da técnica correta.

O presente trabalho teve seu objetivo alcançado uma vez que foi possível encontrar um modelo capaz de prever a probabilidade de pagamento de um cliente inadimplente, através do score criado usando os coeficientes estimados pelo modelo. Sendo que o grande diferencial deste para os estudos anteriores, foi a possibilidade de dividir a resposta estudada, tempo de atraso, em três grupos. Isto só foi possível devido à modelagem de dados por Regressão Logística Polinômica.

O próximo passo agora é implantar o modelo na empresa estudada. Após a implantação, será necessário que se faça uma manutenção constante, através de relatórios gerenciais. Isto permitirá verificar se os objetivos da implantação do modelo estão sendo atendidos e também se a performance das estratégias provenientes dele está de acordo com o esperado. As principais informações a serem gerenciadas referem-se ao resultado obtido por cada faixa de score, comparando-se com as previsões estabelecidas, através da análise das perdas e recuperação para cada categoria de risco.

10. BIBLIOGRAFIAS

AGRESTI, A. (1990) *Categorical Data Analysis*. Series in Probability and Mathematical Statistics. Nova York: John Wiley.

ANGELO, C.F. (coord.), SILVEIRA, J.A.G. (coord.) (1996) *Varejo competitivo*. São Paulo:Atlas,v.2.

ARZULLAH, A.M., KHURI, S.F., HENDERSON, W.G., DALEY, J. *Development and Validation of a Multifactorial Risk Index for Predicting Postoperative Pneumonia after Major Noncardiac Surgery*. Annals of Internal Medicine, v. 135, n. 10, p. 847-857, novembro/2001.

BARROS, A.S. (2003) *Cobrança: a importância da boa gestão para recuperar valores e fidelizar clientes*. Disponível em: www.equifax.com.br

BERNI, M.T. (1999) *Operação e concessão de crédito : os parâmetros para a decisão de crédito*. São Paulo: Atlas.

BITTENCOURT, H.R. *Regressão Logística Polinômica: Revisão teórica e aplicações*. Revista Acta Scientiae, Canoas, 2002

BRAGA, M.C.M. (2003) *Caracterização teórica e aplicação da análise discriminante*. Monografia (Bacharelado em Estatística), Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre.

CAMPOS JÚNIOR, N. (2003) *Sua Excelência, o devedor*. Disponível em: www.equifax.com.br

CAOUILLE, J.B., ALTMAN, E.I., NARAYANAN, P. (1999) *Gestão do Risco de Crédito: O próximo grande desafio financeiro*; tradução de Allan Hastings; revisão técnica João Carlos Douat. Rio de Janeiro: Qualitymark Ed. Tradução de Managing Credit Risk – The Next Financial Challenge.

CARBALLO, M.T. (2002) *Predição da Macrossomia Fetal Através da Regressão Logística e de Redes Neurais Artificiais*. Monografia (Bacharelado em Estatística), Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre.

CHAIA, A.J. (2003) *Modelos de gestão do risco de crédito e sua aplicabilidade ao mercado brasileiro*. Dissertação (Mestrado em Administração), Universidade de São Paulo–USP, São Paulo.

Dicionário Brasileiro Globo, 45 ed., São Paulo: Globo, 1996.

ETTINGER, R.P., GOLIEB, D.E. (1971) *Creditos y Cobranzas*; tradução de Enrique Wallberg; revisão técnica Joaquin Blanes Prieto C.P.T. México: Continental. Tradução de Credits and Collections

FIGUEIREDO, R.P. (2001) *Gestão de Riscos Operacionais em Instituições Financeiras – Uma abordagem qualitativa*. Dissertação (Mestrado em Administração), Universidade da Amazônia – UNAMA, Belém – Pará

GRINBAUM, R. Lojas crescem com cartões. ISTOÉ Dinheiro. Ed 317, 24 Set/2003.

GUIMARÃES, I.A., CHAVES NETO, A. *Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor*. RAE eletrônica, v. 1, n. 2, jul-dez/2002

GUJARATI, D. (2000) *Econometria Básica*. São Paulo: Makron Books

HOSMER, D.W., LEMESHOW, S. (1980) *A goodness-of-fit test for the multiple logistic regression model*. Communications in Statistics, A10, p. 1043-1069.

HOSMER, D.W., LEMESHOW, S. (1989) *Applied Logistic Regression*. Series in Probability and Mathematical Statistics. Nova York: John Wiley, 1989

KLARMANN, P.C. (1993) *Regressão Logística Politémica Nominal e Ordinal*. Monografia (Bacharelado em Estatística), Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre.

MANFIO, F. (2003) *A prática atual de gerenciamento de riscos*. Disponível em: www.equifax.com.br

MARTIN, C.L.M. (1997) *Recuperação de Créditos : Estratégias e Soluções para a Inadimplência*. São Paulo: Érica.

MOHALLEM, W. (2003) *Novo panorama para terceirização de Cobrança*. Disponível em: www.equifax.com.br

PAMPLONA, J. (2004) *A evolução recente de crédito pessoa física*. Disponível em: www.equifax.com.br

PARENTE, J. (2000) *Varejo no Brasil : Gestão e Estratégia*. São Paulo: Atlas.

PEREIRA, A.P.F., BARROSO, M.H., NEPOMUCENO FILHO, F. *Uso do Credit Score na análise de pessoa física*. In: Congresso Nacional de Excelência em Gestão; 2002 Nov 22-23, Niterói, Brasil.

RIBEIRO, T. (2003) *Risco de Crédito*. Risk Control. Disponível em: www.riskcontrol.com.br

RIOS, A.R. (2003) *Cobrança: políticas de recuperação de créditos*. Disponível em: www.equifax.com.br

ROSA, P.T.M. (2000) *Modelos de "Credit Scoring": Regressão Logística, CHAID e REAL*. Dissertação, Universidade de São Paulo – USP, São Paulo.

SALLES-COSTA, R., WERNECK, G.L., LOPES, C.S., FAERSTEIN, E. *Associação entre fatores sócio-demográficos e prática de atividade de lazer no Estudo Pró-Saúde*. Cad. Saúde Pública, Julho/Agosto. 2003, vol.19, no.4, p.1095-1105.

SANTOS, J. O. (2000) *Análise de crédito : empresas e pessoas físicas*. São Paulo: Atlas.

SCHWENCK, P.M. (1977) *Manual prático de crédito e cobrança*. São Paulo: Sugestões Literárias.

SECURATO, J.R. (coord.) (2002) *Análise e Avaliação do Risco – Pessoas Físicas e Jurídicas*. São Paulo: Saint Paul.

SOUSA, A.F., CHAIA, A.J. *Política de crédito: uma análise qualitativa dos processos em empresas*. Caderno de Pesquisas em Administração, São Paulo, julho/setembro 2003, v. 07, n. 3

SOUZA, R.B. (2000) *O modelo de collection scoring como ferramenta para a gestão estratégica do risco de crédito*. Dissertação (Mestrado em Administração), Fundação Getúlio Vargas - FGV, São Paulo.

THOMAS, L.C. (2003) *Novos Avanços das Metodologias de Credit Scoring*. Tecnologia de Crédito, mar/2003.

ANEXOS

Tempo de Relacionamento (anos)		Atraso			Total	BR/MR	BR/AR
		11 a 30 dias	31 a 60 dias	61 a 180 dias			
menos de 1	n	6.174	2.764	1.155	10.093		
	% atraso	25,76	21,63	27,37	24,64	1,19	0,94
	Res. Ajust	6,25	-9,50	4,35			
1	n	2.620	1.260	499	4.379		
	% atraso	10,93	9,86	11,82	10,69	1,11	0,92
	Res. Ajust	1,88	-3,65	2,52			
2	n	1.824	982	314	3.120		
	% atraso	7,61	7,69	7,44	7,62	0,99	1,02
	Res. Ajust	-0,06	0,36	-0,45			
3	n	1.320	733	239	2.292		
	% atraso	5,51	5,74	5,66	5,60	0,96	0,97
	Res. Ajust	-0,92	0,84	0,20			
4	n	1.179	717	230	2.126		
	% atraso	4,92	5,61	5,45	5,19	0,88	0,90
	Res. Ajust	-2,94	2,59	0,81			
5	n	5.910	3.337	925	10.172		
	% atraso	24,66	26,12	21,92	24,83	0,94	1,12
	Res. Ajust	-0,97	4,06	-4,62			
6	n	2.044	1.205	374	3.623		
	% atraso	8,53	9,43	8,86	8,84	0,90	0,96
	Res. Ajust	-2,68	2,82	0,04			
7	n	853	538	167	1.558		
	% atraso	3,56	4,21	3,96	3,80	0,85	0,90
	Res. Ajust	-3,07	2,90	0,55			
8	n	494	287	94	875		
	% atraso	2,06	2,25	2,23	2,14	0,92	0,93
	Res. Ajust	-1,25	1,04	0,43			
9	n	262	161	31	454		
	% atraso	1,09	1,26	0,73	1,11	0,87	1,49
	Res. Ajust	-0,35	1,98	-2,45			
10	n	218	138	38	394		
	% atraso	0,91	1,08	0,90	0,96	0,84	1,01
	Res. Ajust	-1,29	1,65	-0,43			
11	n	34	22	8	64		
	% atraso	0,14	0,17	0,19	0,16	0,82	0,75
	Res. Ajust	-0,88	0,55	0,58			
12	n	21	17	3	41		
	% atraso	0,09	0,13	0,07	0,10		
	Res. Ajust	-0,95	1,42	-0,63			
13	n	119	59	11	189		
	% atraso	0,50	0,46	0,26	0,46	1,08	1,90
	Res. Ajust	1,25	0,01	-2,03			
14	n	93	64	19	176		
	% atraso	0,39	0,50	0,45	0,43	0,77	0,86
	Res. Ajust	-1,53	1,49	0,22			
15	n	126	84	25	235		
	% atraso	0,53	0,66	0,59	0,57	0,80	0,89
	Res. Ajust	-1,53	1,51	0,17			
16	n	220	162	32	414		
	% atraso	0,92	1,27	0,76	1,01	0,72	1,21
	Res. Ajust	-2,23	3,51	-1,73			
17	n	201	117	28	346		
	% atraso	0,84	0,92	0,66	0,84	0,92	1,26
	Res. Ajust	-0,16	1,06	-1,36			

Tempo de Relacionamento (anos)		Atraso			Total	BR/MR	BR/AR
		11 a 30 dias	31 a 60 dias	61 a 180 dias			
18	n	230	115	26	371	1,07	1,56
	% atraso	0,96	0,90	0,62	0,91		
	Res. Ajust	1,37	-0,08	-2,10			
19	n	27	14	2	43		
	% atraso	0,11	0,11	0,05	0,10		
	Res. Ajust	0,57	0,19	-1,22			
Total	n	23.969	12.776	4.220	40.965		
	% atraso	100	100	100	100		