

FACULDADE IBMEC SÃO PAULO
Programa de Mestrado Profissional em Economia

Daniel Evangelista Régis

**APLICAÇÃO DO MODELO MULTI-ESTADO DE MARKOV
EM CARTÕES DE CRÉDITO**

São Paulo
2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Daniel Evangelista Régis

**APLICAÇÃO DO MODELO MULTI-ESTADO DE MARKOV
EM CARTÕES DE CRÉDITO**

Dissertação apresentada ao Programa de Mestrado Profissional em Economia da Faculdade Ibmec São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Economia

Área de concentração: Finanças

Orientador: Prof. Dr. Rinaldo Artes – Ibmec SP

**São Paulo
2007**

Régis, Daniel Evangelista

Aplicação do Modelo Multi-estado de Markov em Cartões de Crédito / Daniel Evangelista Régis; orientador Rinaldo Artes – São Paulo: Ibmec São Paulo, 2007; 58 p.

Dissertação (Mestrado – Programa de Mestrado Profissional em Economia. Área de concentração: Finanças) – Faculdade Ibmec São Paulo.

1.Modelo Multi-estado de Markov 2.Cartões de Crédito 3.Regressão Logística 4.Behaviour Scoring 5.Anti-attrition Scoring

FOLHA DE APROVAÇÃO

Daniel Evangelista Régis
Aplicação do Modelo Multi-estado de Markov em Cartões de Crédito

Dissertação apresentada ao Programa de
Mestrado Profissional em Economia Ibmec
São Paulo, como requisito para obtenção do
título de Mestre em Economia.

Área de concentração: Finanças

Aprovado em: julho/2006

Banca examinadora

Prof. Dr. Rinaldo Artes (orientador)
Instituição: IBMEC SÃO PAULO

Prof^a. Dra. Andrea Maria Accioly Fonseca Minardi
Instituição: IBMEC SÃO PAULO

Prof. Dr. Eduardo Almeida Prado
Instituição: EAESP-FGV

“Vê mais longe a gaivota que voa mais alto”

(Richard Bach)

Agradecimentos

Agradeço ao professor Rinaldo, pelo constante apoio, orientação e por ter acreditado neste projeto desde o início.

Aos meus pais, pelo exemplo de vida e por tudo que me ensinaram.

À minha irmã Aline, por ser uma brava lutadora.

À minha namorada Talitha, por todo o incentivo e por estar comigo em todos os momentos.

Aos amigos Eduardo e Jean, pela constante orientação e por terem acreditado em mim desde o início.

À professora Andrea, pelas sugestões que enriqueceram este trabalho.

Ao colega de turma Rafael, pelo companheirismo e amizade.

Aos colegas de trabalho Osvaldo, Gustavo, Igor, Sandro, Adriano e Tsuyoshi pelas experiências compartilhadas.

À Ângela e à Iracema, que gentilmente forneceram indicações de leitura na área desta dissertação.

Ao Christopher, autor do pacote MSM, por gentilmente responder a dúvidas sobre o software.

À Cheti, pela orientação e incentivo durante a graduação.

Resumo

Régis, Daniel Evangelista. **Aplicação do Modelo Multi-estado de Markov em Cartões de Crédito**. São Paulo, 2006. 58 p. Dissertação (Mestrado) – Faculdade Ibmecc São Paulo, São Paulo, 2006.

Modelos multi-estado de Markov são utilizados na área médica para estimar as probabilidades de transição entre, por exemplo, vários estágios de uma doença, podendo o paciente recuperar-se ou morrer. O principal interesse deste trabalho é analisar a aplicação do modelo multi-estado de Markov na área de risco associado ao uso de cartões de crédito, aproveitando as características de transições entre diversos estados de relacionamento entre os clientes e as instituições ao longo do tempo e, com isso, gerar modelos de score para diversos fins. Modelos de regressão logística também são estimados a fim de comparar os resultados com os obtidos pelo modelo multi-estado de Markov.

Palavras-chave: Modelo Multi-estado de Markov, Cartões de Crédito, Regressão Logística, Behaviour Scoring, Anti-attrition Scoring

Abstract

Régis, Daniel Evangelista. **Application of the Multi-state Markov Model in Credit Cards**. São Paulo, 2006. 58 p. Dissertation (Mastership) – Faculdade Ibmec São Paulo, São Paulo, 2006.

Multi-state Markov models are used in the medical area so as to estimate transition probabilities between, for instance, various states of a disease, in which the patient can recover or die. The main interest of this work is to analyse the application of the Multi-state Markov model in the risk area associated to the use of credit cards, by means of investigating the characteristics of different state transitions of the relationship between clients and institutions throughout time and, therefore, generating score models for diverse purposes. Logistic regression models are also estimated in order to compare its results with the ones obtained by the Multi-state Markov model.

Keywords: Multi-state Markov Model, Credit Cards, Logistic Regression, Behaviour Scoring, Anti-attribution Scoring

Sumário

CAPÍTULO 1 - INTRODUÇÃO	1
1.1 OBJETIVO E MOTIVAÇÃO	1
1.2 O MERCADO DE CARTÕES DE CRÉDITO NO BRASIL	2
1.2.1 História do Cartão de Crédito	2
1.2.2 Participação no sistema	4
1.2.3 Gerenciamento do risco de crédito e relacionamento	5
1.3 MODELOS DE CREDIT SCORING E ANTI-ATTRITION SCORING	6
1.3.1 Modelos de Credit Scoring	6
1.3.2 Modelo de Anti-attrition Scoring	7
CAPÍTULO 2 - METODOLOGIA.....	8
2.1 MODELO DE REGRESSÃO LOGÍSTICA.....	8
2.2 CONCEITOS DE ANÁLISE DE SOBREVIVÊNCIA.....	9
2.2.2 Modelo Semiparamétrico de Cox	11
3.2.3 Modelo Paramétrico	12
2.3 MODELO MULTI-ESTADO DE MARKOV	13
2.3.1 Verossimilhança para o Modelo Multi-estado.....	15
2.3.2 Covariadas	16
2.4 MEDIDAS DE PERFORMANCE	16
2.4.1 Estatística de Kolmogorov-Smirnov	17
2.4.2 Coeficiente de Gini.....	17
2.4.3 Gráficos de Back Test	18
2.5 TRANSIÇÕES EM CARTÕES DE CRÉDITO	19
CAPÍTULO 3 – BASE DE DADOS	22
3.1 BASE DE DADOS.....	22
3.2 TRATAMENTO E SELEÇÃO DE VARIÁVEIS.....	23
3.2.1 Variáveis Explicativas.....	24
CAPÍTULO 4 - RESULTADOS	26
4.1 MODELOS DE REGRESSÃO LOGÍSTICA	26
4.1.1 Modelos de probabilidade de default em 6 e 12 meses	27
4.1.2 Modelos de probabilidade de cancelamento em 6 e 12 meses.....	29
4.2 ESTIMADORES MODELO MULTI-ESTADO DE MARKOV	30
4.2.1 Transição em dia para rotativo (1-2)	33
4.2.2 Transição em dia para em atraso (1-3)	33
4.2.3 Transição em dia para cancelamento voluntário (1-4)	34
4.2.4 Transição rotativo para em dia (2-1)	34
4.2.5 Transição rotativo para em atraso (2-3)	35
4.2.6 Transição rotativo para cancelamento voluntário (2-4)	36
4.2.7 Transição em atraso para em dia (3-1)	36
4.2.8 Transição em atraso para rotativo (3-2)	37
4.2.9 Transição em atraso para cancelamento voluntário (3-4)	37
4.2.10 Transição em atraso para default (3-5)	38
4.2.11 Comentários gerais	38
4.3 COMPARAÇÃO ENTRE O MODELO MULTI-ESTADO DE MARKOV E A REGRESSÃO LOGÍSTICA	39
4.3.1 Indicadores de Desempenho	40

CAPÍTULO 5 - CONCLUSÃO.....	45
APÊNDICE A.....	47
LISTA DE VARIÁVEIS ANALISADAS	47
APÊNDICE B.....	48
EXEMPLOS DE BASES DE DADOS	48
Modelo Multi-estado de Markov	48
Modelo Regressão Logística.....	49
APÊNDICE C	50
BASELINES E TRANSIÇÕES	50
Baselines.....	50
Transições.....	50
APÊNDICE D – PROGRAMA MSM- R	51
REFERÊNCIAS BIBLIOGRÁFICAS	55

Capítulo 1 - Introdução

1.1 Objetivo e Motivação

O mercado de cartões de crédito no Brasil encontra-se em fase de amadurecimento desde a explosão de demanda favorecida pela estabilização da economia brasileira após 1994, com a criação do plano real. A implantação do Sistema de Pagamentos Brasileiro (SPB) em 2002 foi outro fator que impulsionou o mercado de cartões de crédito, através de uma forte migração dos pagamentos em cheque para os pagamentos em meio eletrônico (Figueiredo, 2006).

A estabilidade econômica também influenciou fortemente na maior utilização de sofisticadas metodologias estatísticas no mercado de produtos de crédito como um todo e, em particular, no mercado de cartões de crédito. Uma série de modelos estatísticos vem sendo usada para auxiliar na concessão, no acompanhamento, na cobrança e na retenção de clientes. Regressão logística, análise discriminante, análise de sobrevivência, árvores de decisão, inferência bayesiana e redes neurais são algumas das técnicas utilizadas, dentre as quais, a regressão logística é provavelmente a mais utilizada.

Na regressão logística busca-se estimar a probabilidade de o cliente transitar de um estado A, adimplente, para um estado B, inadimplente, por exemplo, em um determinado período de tempo. Outros tipos de transições não são considerados. Sabe-se que em cartões de crédito existem vários estados possíveis durante o relacionamento entre o cliente e a instituição financeira, tais como *em dia sem utilização de crédito rotativo*, *em dia com utilização de crédito rotativo*, *em atraso*, *cancelamento voluntário* e *default*. Os indivíduos vão assumindo estes estados ao longo do tempo, sendo esta uma característica de eventos recorrentes (Paes, 1999). O fato de existirem vários estados possíveis caracteriza eventos multi-estado (Hougaard, 1999).

A motivação para este trabalho vem do interesse em aproveitar as características de recorrência para gerar matrizes de probabilidades de transição

entre os diversos estados possíveis ao longo do tempo. Para isso será testado o Modelo Multi-estado de Markov (Jackson, 2006), utilizado na área médica para estimar as probabilidades de transição entre, por exemplo, vários estágios de uma doença, podendo o paciente recuperar-se ou morrer (Jackson, 2006). Modelos estimados através da utilização da regressão logística serão desenvolvidos para comparação com alguns dos resultados do modelo multi-estado de Markov.

1.2 O Mercado de Cartões de Crédito no Brasil

O mercado de cartões de crédito no Brasil encontra-se em fase de crescimento acelerado e possui atualmente 85 milhões de cartões (ABECS, 2007). Em 2006, houve um aumento de 17% em quantidade de cartões, 19% em transações realizadas e um aumento da ordem de 23% do faturamento do setor em relação ao ano anterior, passando de R\$ 123 bilhões para R\$ 151 bilhões (ABECS, 2007). Quando comparamos com o ano de 2000 verificamos um crescimento acumulado de quase 200% na quantidade de cartões emitidos, conforme Figura 1.1 (ABECS, 2007).

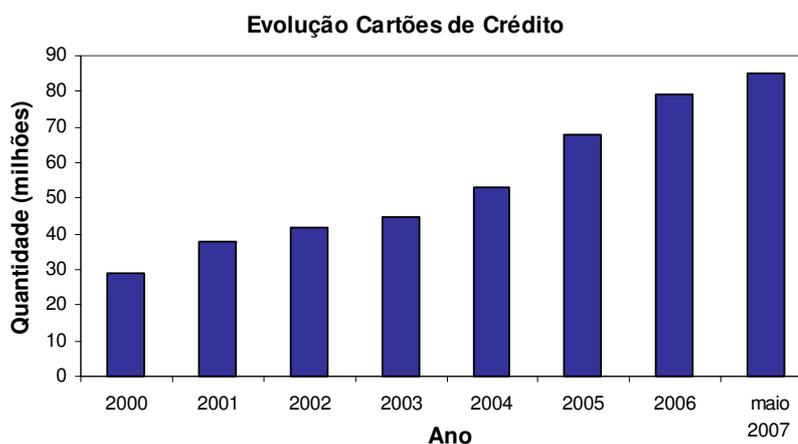


Figura 1.1 – Evolução da quantidade de cartões de crédito (2000 – 2007)

1.2.1 História do Cartão de Crédito

O cartão de crédito surgiu nos Estados Unidos, na década de 20, predominantemente nos postos de gasolina e hotéis, que encontraram nele uma

maneira de prover um tratamento diferenciado e de confiança a seus principais clientes (Portal Economia, 2006).

O Diners Club criou, em 1950, o conceito do cartão de crédito atual, baseado na emissão de cartões por meio de uma empresa estabelecida para esse fim. Após estabelecer uma parceria com um grupo de restaurantes requintados dos Estados Unidos, o Diners Club conseguiu atingir significativa penetração entre os principais executivos do país, que encontravam no cartão uma maneira prática de pagar as despesas de viagem. O produto era confeccionado em papel cartão. O plástico entrou em vigor a partir de 1955 (Portal Economia, 2006).

Em 1958 a American Express, que atuava como agência de viagens, entrou no mercado de cartões, o que passou a despertar a atenção dos bancos. No mesmo ano, o Bank of América foi o pioneiro ao lançar um produto de cartão no segmento bancário (BankAmericard). Em 1977 o BankAmericard passa a denominar-se Visa, que na década de 90 se torna o maior cartão com circulação mundial, sendo aceito em 12 milhões de estabelecimentos. Na atualidade, a maior parte dos cartões é emitida por bancos ou empresas associadas a bancos (Portal Economia, 2006).

No mercado Brasileiro o produto passou a ser utilizado em 1956, mas o investimento propriamente dito só começou a ser estruturado em 1970, com a criação da Credicard, resultado de uma associação entre Citigroup, Itaú e Unibanco, que no final do mesmo ano já contava com 180 mil cartões emitidos e 15 mil estabelecimentos afiliados (Canal Executivo, 2007). Abaixo são citados alguns fatos relevantes:

- 1980: American Express ingressa no mercado brasileiro (Canal Executivo, 2007);
- 1983: Credicard e Visa se associam para a emissão dos cartões desta bandeira, o que as levou à segunda colocação entre os emissores (com 550 mil cartões), atrás apenas do Bradesco (1,2 milhões de cartões) (Canal Executivo, 2007);
- 1984: Credicard adquire o Diners Club (Canal Executivo, 2007);

- 1987: Credicard rompe com a Visa e passa a emitir com exclusividade os cartões com a bandeira Mastercard (Canal Executivo, 2007);
- Início dos anos 90: quebra de exclusividade de bandeiras e estabilização econômica incentiva o surgimento de novos competidores e o surgimento de novos produtos, propiciando um “boom” no mercado de cartões (Canal Executivo, 2007);
- 2004: Unibanco retira-se do negócio da Credicard, ficando a administração a cargo do Itaú e do Citigroup. Nessa época a Credicard já contava com a administração de aproximadamente 7,6 milhões de cartões. É iniciado o processo de cisão da Credicard entre Itaú e Citigroup (Folha Online, 2004).
- 2006: Itaú e Citigroup concluem processo de cisão da Credicard. Itaú passa a ter uma participação de 22% do faturamento do mercado de cartões de crédito com bandeira, liderando o ranking de emissores no Brasil. Bradesco compra operações da empresa de cartões de crédito American Express no Brasil e passa a ter uma participação de 17,6% do faturamento, segundo lugar no ranking de emissores (Clipping Ministério do Planejamento, 2006). Citigroup exerce opção de exclusividade da utilização da marca Credicard a partir de 2009 (Folha Online, 2006). Outros dois grandes concorrentes neste mercado são Banco do Brasil e Unibanco.

1.2.2 Participação no sistema

Os cartões de crédito possuem emissor, bandeira e variante. Abaixo estão listados os papéis dos agentes desse mercado:

Emissores: são Instituições Financeiras que emitem e administram cartões próprios ou de terceiros e concedem financiamento direto aos portadores ou a Administradoras, instituições não financeiras que emitem e administram cartões próprios ou de terceiros, mas não financiam diretamente os seus clientes.

Acquirer (adquirente): empresas responsáveis pela filiação, gerenciamento e relacionamento com os estabelecimentos comerciais e pelas condições comerciais. Os acquires em operação no Brasil são: Redecard, Visanet e Amex.

Bandeiras: são instituições que autorizam o uso de sua marca e de sua tecnologia por emissores e acquirers. As principais bandeiras do mercado brasileiro são Visa, MasterCard, Diners Club, Redeshop e American Express.

Processadoras: são empresas que prestam serviços operacionais relacionados à administração de cartões. As principais marcas do mercado brasileiro são Orbitall, Cardsystem e Equifax/Unnisa

Variantes: limites de crédito diferenciados, dentro de um mesmo emissor, associados à renda do cliente.

Parceiros: empresas que atuam em outra área e fazem acordo com os emissores para oferecer um cartão com sua marca. Geralmente oferecem algum diferencial (serviço, produto ou facilidade). Os parceiros podem ser redes varejistas, companhias aéreas ou automobilísticas, entidades beneficentes ou indústrias.

1.2.3 Gerenciamento do risco de crédito e relacionamento

Com o crescimento do mercado de cartões de crédito é natural que ocorra uma maior preocupação com os níveis de inadimplência nas instituições. Modelos de concessão (application scoring) são amplamente utilizados para avaliar o risco de crédito de uma nova operação e modelos de acompanhamento (behaviour scoring) são utilizados no gerenciamento do risco de crédito de clientes que já possuem algum produto.

Manter um longo relacionamento com clientes de cartões de crédito também é de fundamental importância para as instituições, uma vez que estima-se que obter um novo cliente é até dez vezes mais caro do que reter um cliente existente

(Oliveira, 2000). Modelos de acompanhamento do relacionamento (anti-attrition scoring) são utilizados no auxílio à retenção de clientes.

1.3 Modelos de Credit Scoring e Anti-attrition Scoring

1.3.1 Modelos de Credit Scoring

Os modelos de Credit Scoring são sistemas que atribuem pontuações às variáveis de decisão de crédito de um proponente, mediante a aplicação de técnicas estatísticas. Esses modelos visam à segregação de características que permitam distinguir os bons dos maus créditos (Lewis, 1992). O desenvolvimento de modelos de Credit Scoring requer a utilização de técnicas estatísticas avançadas, tais como regressão logística, análise discriminante, árvores de decisão, inferência *bayesiana* e redes neurais, além de conhecimento prático do tipo de cliente a ser analisado.

Os modelos de Credit Scoring podem ser divididos em duas categorias: modelos de concessão (application scoring) e modelos comportamentais (behaviour scoring). Modelos de application scoring são utilizados para auxiliar instituições financeiras na tomada de decisão de concessão de crédito a um novo cliente. Tais modelos buscam, baseados em características do proponente e da operação de crédito, estimar a probabilidade de inadimplência em um determinado período e utilizam, principalmente, informações cadastrais dos clientes. Em Rosa (2000) e Thomas et al. (2002) são descritas todas as etapas necessárias ao desenvolvimento de um modelo de application scoring.

Modelos de behaviour scoring auxiliam a instituição no gerenciamento do relacionamento com os clientes que já possuem algum produto, sendo utilizados como importante ferramenta nas decisões de manutenção de limites e oferta de novos produtos. Os modelos de behaviour scoring são baseados, principalmente, em características de compra ou pagamento do cliente e por isso apresentam poder de discriminação bastante superior aos observados em modelos de application scoring. Hoper e Lewis (1992) descrevem como um modelo de behaviour scoring é

geralmente utilizado. Blackwell e Sykes (1992) descrevem como modelos de behaviour scoring podem ser utilizados para a decisão de qual o limite de crédito a ser atribuído ao cliente. Ohtoshi (2003) compara a performance de várias metodologias utilizadas no desenvolvimento de modelos de behaviour scoring. Tomas et al. (2001) descreve como criar modelos de behaviour scoring utilizando cadeias de Markov onde o cliente é classificado em um estado de acordo com algumas variáveis e então estima-se a probabilidade de o cliente ir a um estado de default (inadimplência).

Atualmente, tanto modelos de application scoring, quanto modelos de behaviour scoring têm obtido ganhos significativos de performance através da utilização de informações de bureaus de crédito como Serasa e SPC. Nesses modelos, além das informações disponíveis sobre os clientes dentro da instituição, são utilizadas informações do comportamento do cliente no mercado como um todo.

Modelos de credit scoring costumam ser utilizados para dividir a carteira em classes de score. Essas classes definem grupos de clientes com nível semelhante de risco, permitindo à instituição o desenvolvimento de políticas específicas para cada grupo.

1.3.2 Modelo de Anti-attrition Scoring

Modelos de anti-attrition scoring, baseados principalmente em informações de relacionamento e utilização do produto, têm por objetivo identificar antecipadamente clientes com alto potencial de ruptura no relacionamento com a instituição e, com isso, permitir a tomada de ações preventivas para evitar o cancelamento do produto. As metodologias utilizadas para o desenvolvimento de modelos de anti-attrition scoring são basicamente as mesmas utilizadas para o desenvolvimento de modelos de credit scoring.

Capítulo 2 - Metodologia

Nesta dissertação, desejamos prever o estado no qual se encontrará um possuidor de cartão de crédito, a partir de seu estado atual e de informações sobre o seu perfil. Apresentamos, neste capítulo, duas abordagens teóricas para o problema: a regressão logística e o modelo multi-estado de Markov.

2.1 Modelo de Regressão Logística

A regressão logística (ver Hosmer e Lemeshow, 2000, por exemplo), técnica estatística utilizada para estudar a relação entre uma variável categorizada, em geral binária, e um conjunto de outras variáveis, tem sido amplamente utilizada na área de modelagem estatística de dados e, em particular, na área de gestão de risco de crédito. Na regressão logística o interesse é estimar a probabilidade de uma observação assumir determinado estado após um determinado período de tempo. Assumimos que esse estado é absorvente, ou seja, uma vez neste estado, o cliente nele permanece indefinidamente.

Num primeiro momento, consideramos a regressão logística múltipla com resposta binária, tendo como interesse a atribuição de um escore de default e um escore de cancelamento para cada indivíduo da amostra.

Admita que y é uma variável indicadora que assume o valor um quando ocorre um evento de interesse e o valor zero quando esse evento não ocorre. Na regressão logística buscamos encontrar uma relação funcional entre $E(Y)=P(Y=1)$ e X_1, X_2, \dots, X_p , variáveis independentes, modelando-se o logaritmo neperiano da razão de chances (*odds ratio*) de ocorrência do evento de interesse

$$\ln\left(\frac{P(Y_i = 1)}{P(Y_i = 0)}\right) = \alpha + \beta^T x_i, \quad (2.1)$$

em que α é o intercepto, β é um vetor de parâmetros $(\beta_1, \beta_2, \dots, \beta_p)^T$ e x_i é um vetor contendo as variáveis preditoras do modelo para o i -ésimo elemento da amostra: $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$.

A partir de (2.1) temos que

$$P(Y_i = 1) = \frac{e^{\alpha + \beta^T x_i}}{1 + e^{\alpha + \beta^T x_i}}.$$

O método mais eficiente para estimar os parâmetros do modelo de regressão logística é o da máxima verossimilhança, onde a função de verossimilhança, mostrada em (2.2) é obtida a partir da probabilidade de reproduzir os dados da amostra segundo o modelo.

$$L(\beta) = \prod_{i=1}^n [P(Y_i = 1)]^{y_i} [1 - P(Y_i = 1)]^{1-y_i}, \quad (2.2)$$

2.2 Conceitos de Análise de Sobrevida

A análise de sobrevivência (ver Colosimo e Giolo, 2006, por exemplo), base do desenvolvimento do modelo multi-estado de Markov, é um conjunto de procedimentos estatísticos para a modelagem de dados relacionados ao tempo até a ocorrência de um determinado evento de interesse. Na análise de sobrevivência, o tempo até a ocorrência de determinado evento, chamado de tempo de falha, é a variável de interesse. Por diversos motivos muitas vezes o tempo de falha não é conhecido, o que caracteriza censura, ou seja, a observação parcial da resposta. A seguir são definidos alguns conceitos de análise de sobrevivência, segundo Colosimo e Giolo (2006):

Tempo de falha

Tempo até a ocorrência do evento de interesse. Representado pela variável aleatória T .

Censura

Presença de observações incompletas ou parciais devido ao término do estudo antes que todos os indivíduos venham a falhar. Mesmo censurados, todos os resultados de um estudo de sobrevivência devem ser usados na análise, para evitar conclusões viciadas e para fornecer informações sobre o tempo de falha, mesmo que incompletas.

Função de sobrevivência

Probabilidade de uma observação não falhar até um certo tempo t , ou seja,

$$S(t) = P(T > t)$$

Função de distribuição acumulada

É a probabilidade de uma observação falhar até o tempo t , ou seja,

$$F(t) = P(T \leq t) = 1 - S(t)$$

Função de risco ou taxa de falha:

A probabilidade de a falha ocorrer em um intervalo de tempo $[t_1, t_2)$ é:

$$S(t_1) - S(t_2)$$

Temos que a probabilidade de haver uma falha até o tempo t_2 , uma vez que não houve falha até o tempo t_1 é dada por

$$P(t_1 \leq T < t_2 | T > t_1) = \frac{S(t_1) - S(t_2)}{S(t_1)}.$$

Define-se a taxa de falhas no intervalo $[t_1, t_2)$ como

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} = \frac{P(t_1 \leq T < t_2 | T > t_1)}{(t_2 - t_1)} \quad (2.3)$$

Redefinindo o intervalo como $[t, t+\Delta t)$ reescrevemos (2.3) como

$$\frac{(S(t) - S(t + \Delta t))}{\Delta t S(t)}.$$

Fazendo Δt tendendo a zero, $\lambda(t)$, definida em (2.4), é denominada como a taxa de falha instantânea no tempo t condicional à sobrevivência nesse tempo.

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}, \quad (2.4)$$

Prova-se que a função de risco pode também ser escrita em termos da função sobrevivência e da função de densidade de T ($f(t)$) (ver, detalhes em Colosimo e Giolo, 2006), conforme (2.5).

$$\lambda(t) = \frac{f(t)}{S(t)}, \quad (2.5)$$

Existem três tipos de modelos para a estimação das funções de sobrevivência: modelos não-paramétricos, modelos semiparamétricos e modelos paramétricos. Descreveremos brevemente o modelo semiparamétrico de Cox e o modelo paramétrico.

2.2.2 Modelo Semiparamétrico de Cox

O modelo semiparamétrico de Cox, ou de riscos proporcionais, possibilita a inclusão de efeitos de características individuais na probabilidade de ocorrer falha, sendo tais características utilizadas como variáveis explicativas, ou covariáveis da variável resposta.

A função risco, atribuída ao elemento i da amostra, é dada por

$$\lambda(t) = \lambda_0(t) \exp(x_i^T \beta)$$

sendo λ_0 uma função não negativa denominada função de base. Note que $\lambda(t) = \lambda_0(t)$ quando $x = 0$, onde x é um vetor de covariáveis definido na Seção 2.1. O modelo para a função risco é chamado semiparamétrico porque apenas os efeitos das covariáveis são tratados parametricamente.

Para a estimação dos parâmetros é utilizado o método de máxima verossimilhança parcial, obtido maximizando-se a função de verossimilhança parcial dada abaixo:

$$L(\beta) = \prod_{i=1}^n \left(\frac{\exp(x_i^T \beta)}{\sum_{j \in R(t_i)} \exp(x_j^T \beta)} \right)^{\delta_i},$$

onde $R(t_i)$ é o conjunto dos índices das observações sob risco no tempo i e δ_i é o índice de falha.

3.2.3 Modelo Paramétrico

Para a utilização de métodos paramétricos para a modelagem de tempo até o evento de interesse, em função de covariáveis, é necessária a especificação de uma distribuição de probabilidade para a variável resposta. Distribuições de probabilidade geralmente utilizadas para descrever o tempo até a ocorrência de determinado evento são a exponencial, a Weibull e a log-normal. A distribuição exponencial é, segundo Colosimo e Giolo (2006), um dos modelos mais simples e historicamente mais utilizado na literatura de análise de sobrevivência. Para a estimação dos parâmetros, o método de máxima verossimilhança, utilizando a função $L(\beta)$ abaixo, é apropriado.

$$L(\beta) = \prod_{i=1}^r f(t_i; \beta) \prod_{i=r+1}^n S(t_i; \beta)$$

Na função $L(\beta)$ as observações são divididas em dois conjuntos, as r primeiras são as não-censuradas e as $(n-r)$ seguintes são as censuradas.

2.3 Modelo Multi-estado de Markov

Estamos admitindo uma situação em que os possuidores de cartão de crédito podem assumir diferentes estados (por exemplo: *em dia*, *em atraso*, *em default*) ao longo do tempo. Alguns desses estados são transitórios, ou seja, o cliente pode sair dele em algum momento (por exemplo, um cliente em atraso pode voltar a ficar em dia) e outros absorventes, ou seja, uma vez nesse estado, o cliente não sai mais dele (por exemplo um cliente em *default* sai da base de clientes e não pode mais assumir outros estados). O modelo multi-estado de Markov assume que as probabilidades de transições entre estados depende apenas do tempo entre as transições e de covariáveis (eventualmente dependentes do tempo) associadas aos clientes. Detalhes sobre o desenvolvimento teórico desses modelos podem ser encontrados em Kalbfleisch e Lawless (1985), Kay (1986), Jackson et al. (2003) e Jackson (2006), por exemplo.

As Figuras 2.1 e 2.2 representam situações em que são permitidas mudanças de estados ao longo do tempo. A Figura 2.1 representa uma situação em que existem três estados transitórios. As setas indicam que é possível passar de um estado ao outro diretamente. A Figura 2.2, por sua vez, é uma reprodução de um estudo apresentado em Jackson e Sharpless (2002) e Jackson et al. (2003). O esquema representa a evolução de uma doença. Os estágios 1 até (K-1) são transitórios e indicam a gravidade da doença (quanto maior o número, mais grave é o estado). Já o estado K indica a morte do paciente, sendo obviamente um estado absorvente. O paciente pode avançar para estágios mais graves da doença, retroceder para estágios menos graves, ou morrer a partir de qualquer estágio. É interessante notar que nem todas as transições entre estados transitórios são possíveis diretamente. Por exemplo, para ir do Estado 1 ao Estado 3, é necessário passar, em algum momento pelo Estado 2.

Seja $E_i(t)$, o estado assumido pelo indivíduo i , $i = 1, \dots, n$, no instante $t = 1, \dots, \tau$. Admita a existência de K possíveis estados. A probabilidade de um indivíduo i passar do estado r para o estado s num intervalo de tempo Δt é dada por

$$p_{irs}(\Delta t) = P(E_i(t + \Delta t) = s \mid E_i(t) = r). \quad (2.6)$$

Define-se a intensidade de transição entre os estados r e s por

$$q_{irs}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(E_i(t + \Delta t) = s \mid E_i(t) = r)}{\Delta t} \quad (2.7)$$

Podemos observar que a intensidade de transição, (2.7), assemelha-se a $\lambda(t)$ dada em (2.4), que é a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t . É possível interpretarmos a intensidade de transição como o risco instantâneo de o indivíduo migrar para o estado s a partir do estado r , onde a falha, no caso, é a migração para o estado s .

A matriz $Q_i = [q_{irs}]_{K \times K}$ é denominada matriz de intensidade, e indica as possíveis transições de estado (vide Figuras 2.1 e 2.2). Por conveniência, define-se $q_{irr} = -\sum_{s \neq r} q_{irs}$, exceto se o estado for absorvente, caso em que $q_{irr} = 0$. Desse modo é possível mostrar¹ que as probabilidades de transição (2.6) são obtidas através das componentes da matriz $P_i(\Delta t) = \exp(Q_i \Delta t)$.

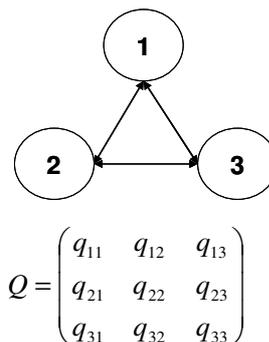


Figura 2.1 – Um modelo multi-estado e matriz de intensidades de transição

¹ Ver detalhes nas referências desta seção e em Cox e Miller (1965).

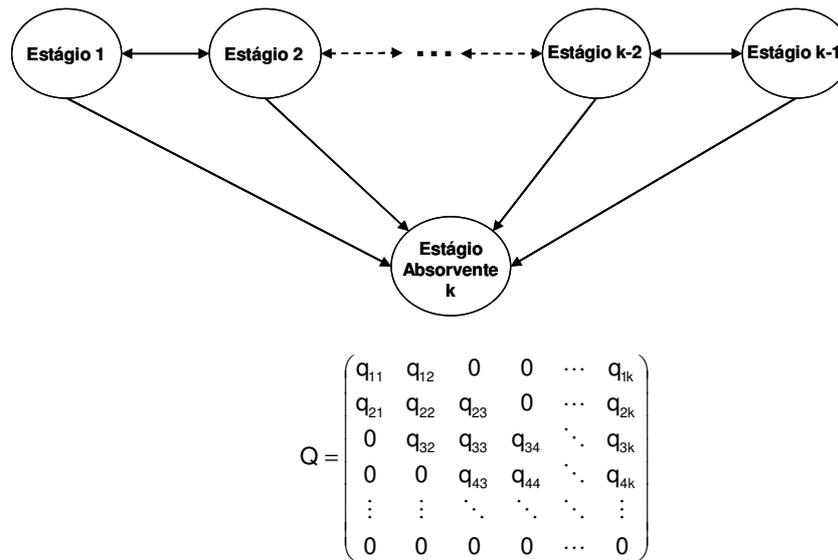


Figura 2.2 – Um modelo geral para progressão de doenças e matriz de intensidades de transição

2.3.1 Verossimilhança para o Modelo Multi-estado

Kalbfleisch e Lawless (1985) e Kay (1986) descrevem um método geral para calcular a verossimilhança para um modelo multi-estado em tempo contínuo, aplicável para qualquer forma de matriz de transição.

A verossimilhança é calculada a partir da matriz de probabilidades de transição $P(t)$. Para um processo homogêneo no tempo, o elemento (r,s) de $P(t)$ é a probabilidade de estar no estado s no tempo $t + u$ no futuro, dado que o estado no tempo t é r .

A série de tempos $(t_{i1}, t_{i2}, \dots, t_{i\tau_i})$ e os correspondentes estados $(E_i(t_{i1}), E_i(t_{i2}), \dots, E_i(t_{i\tau_i}))$ são os dados para o indivíduo i . Considere um modelo multi-estado geral, com um par de sucessivos estados observados da doença $E(t_j)$, $E(t_{j+1})$ nos tempos t_j , t_{j+1} . A contribuição deste par de estados para a verossimilhança é:

$$L_{i,j} = p_{E(t_j)E(t_{j+1})}(t_{j+1} - t_j),$$

onde $L_{i,j}$ é o elemento da matriz de transição $P(t)$ na linha $E(t_j)$ e na coluna $E(t_{j+1})$, calculados em $t = t_{j+1} - t_j$.

A verossimilhança completa $L(Q)$ é o produto de todos os termos $L_{i,j}$ sobre todos os indivíduos e todas as transições e depende da matriz desconhecida de transição Q , que é usada para determinar $P(t)$.

2.3.2 Covariadas

Características constantes ou que variem no tempo dos indivíduos podem ser utilizadas como variáveis explicativas em um modelo multi-estado. Marshall e Jones (1995) descrevem uma forma de modelo de *riscos proporcionais* (semelhante ao modelo semiparamétrico de Cox) na qual os elementos q_{irs} da matriz de intensidade de transições podem ser substituídos por

$$q_{irs}(x_i(t)) = q_{rs}^{(0)} \exp(\beta_{rs}^T x_i(t)), \quad (2.8)$$

sendo $x_i(t)$ o vetor p -dimensional com os valores das variáveis explicativas observadas no instante t para o indivíduo i .

A nova matriz Q_i é então utilizada para determinar a verossimilhança. Se as covariadas $x_i(t)$ são dependentes do tempo, as contribuições para a verossimilhança $p_{rs}(t - u)$ são substituídas por $p_{rs}(t - u, x_i(u))$. Entretanto, este procedimento requer que o valor das covariadas seja conhecido em cada período de observação u . Marshall e Jones (1995) descrevem testes de Wald para a seleção de covariadas e outros testes de hipóteses.

2.4 Medidas de Performance

Existem várias medidas que permitem mensurar e comparar o desempenho de modelos de credit, behaviour e anti-attrition scoring. Neste trabalho serão utilizadas algumas das medidas mais populares como a estatística de Kolmogorov-Smirnov, o coeficiente de Gini e gráficos para inspeção visual.

2.4.1 Estatística de Kolmogorov-Smirnov

Utiliza-se a estatística de Kolmogorov-Smirnov (KS) para comparar as funções de distribuição de escores, denotados por s , entre os clientes bons e maus. Em um bom modelo são atribuídos, predominantemente, altos escores aos bons clientes e baixos escores aos maus clientes. Definindo-se $F_b(s)$ como a frequência relativa acumulada do escore dos clientes bons e $F_m(s)$ a frequência relativa acumulada do escore dos maus clientes, espera-se que $F_m(s)$ se aproxime mais rapidamente de 1 do que $F_b(s)$. Quanto mais rápido o crescimento de $F_m(s)$ e mais lento o crescimento de $F_b(s)$ melhor é o modelo. A estatística de Kolmogorov-Smirnov é definida como a maior distância entre $F_m(s)$ e $F_b(s)$ (Figura 2.3), variando de 0 a 1 e quanto mais próximo de 1 melhor a performance do modelo.

$$KS = \max_s \{F_m(s) - F_b(s)\}$$

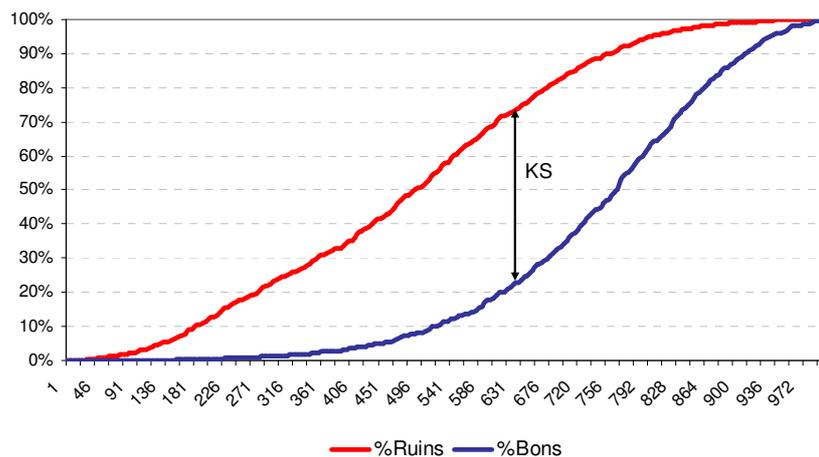


Figura 2.3 – Exemplo de Gráfico de KS

2.4.2 Coeficiente de Gini

O coeficiente de Gini, determinado a partir da curva ROC (Thomas et al., 2002), é baseado nos conceitos de sensibilidade e especificidade, estatísticas que podem ser obtidas a partir da construção de diversas matrizes de confusão (Johnson e Wichern, 2002). Seja c um ponto de corte e admita que um indivíduo será classificado como bom se seu escore for superior a c e como ruim se for inferior. Define-se como sensibilidade associada ao ponto de corte c como a proporção de clientes bons que são classificados corretamente quando se utiliza c como ponto de

corde. Por outro lado, define-se a especificidade associada a c como a proporção de clientes maus que são classificados corretamente através do uso de c como ponto de corte. Quanto maior a sensibilidade e a especificidade, melhor o modelo. Para a obtenção da curva ROC calculam-se, para m pontos de corte distintos, m especificidades e m sensibilidades. A curva ROC é formada pela união dos pontos $(1 - \text{especificidade}, \text{sensibilidade})$ de todos os pares sensibilidade/especificidade obtidos. O coeficiente de Gini é dado por duas vezes a área entre a curva ROC e a reta $Y = X$, variando de 0 a 1, sendo que quanto mais próximo de 1 melhor é o modelo. A Figura 2.4 ilustra a construção de uma curva ROC.

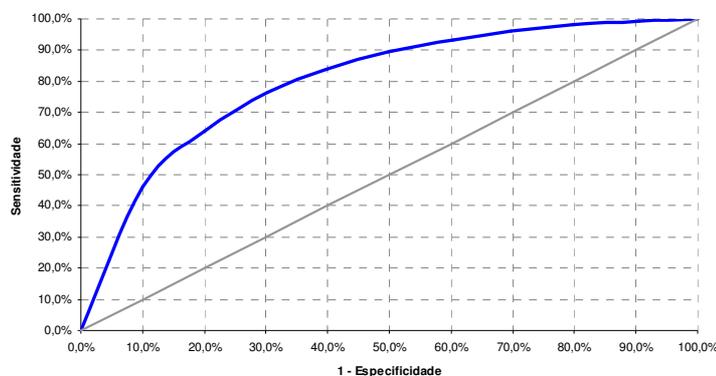


Figura 2.4 – Exemplo de curva ROC

2.4.3 Gráficos de Back Test

Uma técnica simples, mas de elevada eficiência na avaliação de um modelo de credit, behaviour ou anti-attrition scoring, é verificar graficamente a ordenação do escore em relação à variável resposta. Divide-se a população em percentis do escore previsto pelo modelo e verifica-se o percentual de clientes maus em cada percentil. Espera-se que um bom modelo apresente um maior percentual de clientes maus nos percentis de menor escore e um pequeno percentual de clientes maus nos percentis de maior escore, com o percentual de inadimplentes decrescendo monotonicamente da faixa mais baixa para a faixa mais alta, conforme a Figura 2.5.

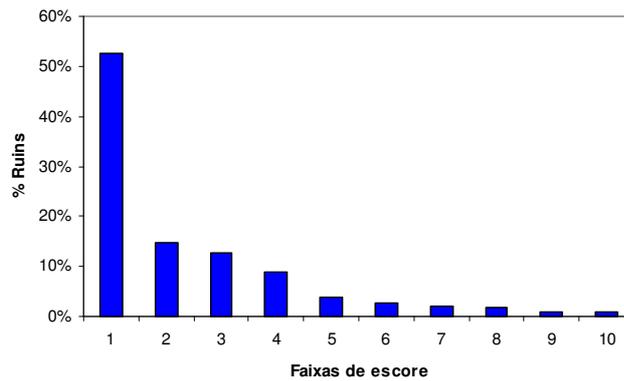


Figura 2.5 – Exemplo de gráfico de back test

2.5 Transições em Cartões de Crédito

Segundo Paes (1999), em muitos estudos médicos que envolvem o acompanhamento de pacientes, alguns eventos de interesse são não terminais e podem ocorrer mais de uma vez para o mesmo indivíduo (eventos recorrentes), o que é uma situação comum em estudos de doenças crônicas, como diabetes, cujo interesse pode ser a ocorrência de hiperglicemias. Alguns outros exemplos são crises de asma, ataques epiléticos, episódios de rejeição em pacientes transplantados, internações hospitalares e infecções de vários tipos. Em casos mais simples observa-se a ocorrência de um único evento, deixando o paciente de ser acompanhado a partir do momento que o evento ocorre. Em eventos recorrentes, o paciente pode continuar em observação mesmo depois da ocorrência de cada evento. Ainda segundo Paes (1999), muitos pesquisadores consideram como período de observação apenas o tempo até o instante em que o evento ocorre pela primeira vez, devido à incapacidade das técnicas usuais, resultando em um desperdício de informações obtidas durante o restante do acompanhamento.

Além dos eventos recorrentes, existem os eventos absorventes que são eventos terminais (estados absorventes) e uma vez estando neles não há como sair, como por exemplo, a morte de pacientes.

As características de eventos recorrentes e absorventes podem ser observadas em outras áreas, além da área médica. Em cartões de crédito, default

pode ser visto como evento absorvente, enquanto que o atraso no pagamento da fatura pode ser considerado um evento recorrente.

Neste trabalho serão analisadas as transições entre os seguintes estados ao longo do tempo:

Estados com características de recorrência:

1 - Em dia (OK): o cliente pagou o total da fatura do mês

2 - Rotativo (R): o cliente pagou parte da fatura do mês, ou seja, algum valor entre o pagamento mínimo e o total da fatura. Neste caso, o cliente encontra-se em situação de regularidade junto à Administradora.

3 - Em atraso (A): o cliente não pagou o valor mínimo da fatura no mês.

Estados absorventes:

4 - Cancelamento voluntário (C): cancelamento do cartão de crédito por iniciativa do cliente.

5 - Default (D): cancelamento do cartão de crédito por iniciativa da Administradora, devido à inadimplência. Neste estudo a ocorrência de 3 atrasos consecutivos caracteriza default.

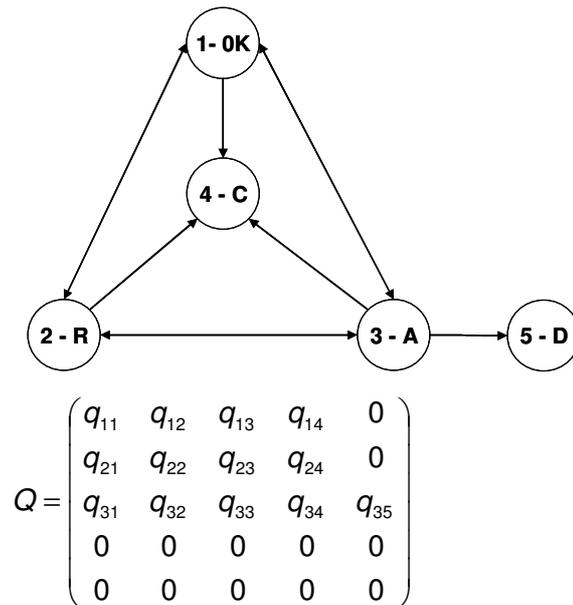


Figura 2.6 – Modelo multi-estado em cartões de crédito e matriz de intensidades de transição

Na Figura 2.6 as possíveis transições e a estrutura da matriz de intensidades são ilustradas. Clientes em dia são representados com o texto OK. As transições possíveis para clientes que se encontram no estado *em dia* são: *em dia* para *rotativo*, *em dia* para *em atraso* e *em dia* para *cancelamento voluntário*. Não é possível um cliente no estado *em dia* migrar para o estado *default* diretamente, uma vez que só é possível, neste estudo, ir a *default* a partir do estado *em atraso*.

O estado *rotativo* está representado pela letra R e as transições possíveis a partir deste estado são: *rotativo* para *em dia*, *rotativo* para *em atraso* e *rotativo* para *cancelado*. O estado *em atraso* está representado pela letra A. As transições possíveis para clientes que se encontram em atraso são: *em atraso* para *em dia*, *em atraso* para *rotativo*, *em atraso* para *cancelamento voluntário* e *em atraso* para *default*. Uma observação importante para as transições a partir do estado em atraso é a de que o cliente, não pode permanecer no estado em atraso por mais de dois meses consecutivos, pois isto caracteriza *default* neste estudo. Para clientes no estado *default*, a única possibilidade é permanecer neste estado em todos os meses seguintes, pois este evento é absorvente. O mesmo vale para clientes no estado *cancelamento voluntário*.

Capítulo 3 – Base de Dados

3.1 Base de Dados

Os dados analisados neste estudo são provenientes de uma grande instituição financeira brasileira que atua no mercado de cartões de crédito. Por motivos de sigilo foi gerada uma carteira artificial, que não reflete os índices de utilização do crédito rotativo, atraso, cancelamento e default da instituição da qual foi retirada.

Para a aplicação do modelo multi-estado de Markov foi gerada uma amostra contendo o histórico de 19 mil indivíduos, sendo 10 mil para desenvolvimento do modelo e 9 mil para validação, selecionados da seguinte forma:

- Pertencer a uma das safras de concessão compreendidas entre os meses de janeiro de 2003 e junho de 2004
- Possuir cartão de crédito ativo em janeiro de 2005, com pelo menos 6 meses de ativação
- Estar em janeiro de 2005 em um dos seguintes estados: *em dia*, *rotativo*, *em atraso*

O objetivo da seleção da amostra acima citada foi obter uma base com um extrato da população que possuía entre 6 e 18 meses de relacionamento com a instituição e que estivesse de fato utilizando o produto. Clientes mais antigos poderiam ser selecionados, mas sem um critério de corte por tempo de relacionamento o estudo das transições poderia ser prejudicado. Clientes muito antigos, por exemplo, podem tratar-se de “*clientes fidelizados*” (ver Quidim, 2005, por exemplo), ou seja, têm uma menor probabilidade de cancelar o cartão por conta própria ou de entrar em default. Poderíamos desenvolver modelos para outros extratos da população a fim de capturarmos os efeitos das variáveis preditoras para clientes com diversos níveis de tempo de relacionamento.

3.2 Tratamento e seleção de variáveis

Para o desenvolvimento do modelo optou-se por uma quantidade reduzida de variáveis comportamentais que indicasse uma forte relação com a tendência de o cliente apresentar problemas de cancelamento, atraso ou default, além da capacidade de discriminação da tendência de utilização do crédito rotativo.

As variáveis históricas construídas basearam-se principalmente em: média aritmética em 12 meses, média exponencial em 6 meses², quantidade de meses com utilização ou em determinado estado, quantidade consecutiva de meses e máxima utilização ou máxima quantidade, aplicadas no comportamento de compras do cliente, no percentual de utilização do limite e no perfil de atraso ou utilização de crédito rotativo.

Para a criação das variáveis históricas, foram utilizados os 12 meses anteriores a janeiro de 2005 e a observação das transições futuras a este mês é feita nos 12 meses seguintes, conforme Figura 3.1. Um exemplo da base utilizada no modelo multi-estado de Markov encontra-se no Apêndice B.

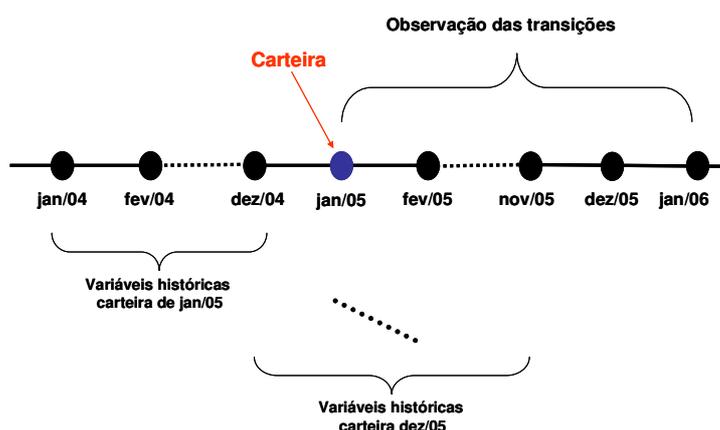


Figura 3.1 – Estrutura da base para o modelo multi-estado de Markov

A base de dados possui 28 variáveis preditivas, listadas no Apêndice A, das quais 5 são cadastrais referentes à operação e 23 são construídas, baseadas no comportamento histórico do cliente no produto. Foram selecionadas 7 variáveis (6

² Trata-se de uma média ponderada em que maiores pesos são dados às observações mais recentes. Maiores detalhes podem ser encontrados em Coelho (2006).

comportamentais e 1 cadastral) utilizando critérios julgamentais e o algoritmo CHAID (*Chi-squared Automatic Interaction Detection* - Rosa, 2000), tendo como variável resposta o estado do cliente após 12 meses do mês de observação inicial (cancelado, default, outros). As variáveis contínuas selecionadas foram agrupadas individualmente em 10 categorias, via CHAID, para reduzir a influência de pontos aberrantes na estimação dos modelos.

3.2.1 Variáveis Explicativas

Variante: variável categorizada em que cada categoria é associada a um determinado nível de limite de crédito, de acordo com a renda do cliente. Foram consideradas 3 categorias de variantes.

Máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses: variável indicadora do perfil de utilização do crédito rotativo, que é uma dos estados possíveis para as transições.

Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses: variável fortemente relacionada à propensão de cancelamento do cliente.

Média exponencial de meses com status atrasado nos últimos 6 meses: variável fortemente relacionada à tendência de um cliente apresentar problemas futuros de atraso ou default.

Média exponencial da quantidade de compras nos últimos 6 meses: variável fortemente relacionada à propensão de cancelamento do relacionamento por parte do cliente.

Média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses: variável fortemente relacionada à propensão de cancelamento por parte do cliente, quando em níveis muito baixos, e à tendência de apresentação de problemas futuros, quando em níveis excessivamente altos.

Maior percentual de utilização do limite de crédito nos últimos 12 meses: variável fortemente relacionada à propensão de cancelamento por parte do cliente, quando em níveis muito baixos, e à tendência de apresentação de problemas futuros, quando em níveis altos.

Além das variáveis selecionadas, são necessárias as seguintes variáveis para o modelo multi-estado de Markov:

ID: identificação do registro, associa cada cliente observado a um número.

Tempo: tempo decorrido desde o início da análise das transições (tempo=0 em janeiro de 2005 e tempo=12 em janeiro de 2006).

Status: estado no qual se encontra o cartão de crédito do cliente, em cada período de tempo.

Para os modelos de regressão logística, além das variáveis selecionadas são necessárias as seguintes variáveis:

ID: identificação do registro.

Performance: variável binária que indica, dependendo do modelo, se o evento *cancelamento voluntário* ou *default* aconteceu durante os 6 ou 12 meses após o mês de observação.

Capítulo 4 - Resultados

4.1 Modelos de Regressão Logística

Para efeito de comparação de alguns dos resultados do modelo multi-estado de Markov com a regressão logística múltipla com resposta binária, foram desenvolvidos 4 modelos distintos:

Modelo L1: Modelo de regressão logística com a finalidade de estimar a probabilidade de o cliente entrar em default no cartão de crédito em um horizonte de 6 meses.

Modelo L2: Modelo de regressão logística com a finalidade de estimar a probabilidade de o cliente entrar em default no cartão de crédito em um horizonte de 12 meses.

Modelo L3: Modelo de regressão logística com a finalidade de estimar a probabilidade de cancelamento do cartão de crédito em um horizonte de 6 meses.

Modelo L4: Modelo de regressão logística com a finalidade de estimar a probabilidade de cancelamento do cartão de crédito em um horizonte de 12 meses.

As mesmas variáveis independentes utilizadas no modelo multi-estado de Markov foram consideradas. As variáveis de comportamento, agrupadas em 10 faixas cada uma, foram tratadas de forma contínua e a variável *variante* foi tratada de forma categorizada. Utilizando-se a técnica conhecida como stepwise (Paula, 2004), foram eliminadas algumas variáveis não significativas em cada modelo.

O esquema mostrado na Figura 4.1 ilustra a estrutura das bases utilizadas para os modelos de regressão logística com variável resposta observada após um horizonte de 12 meses. Um exemplo da base utilizada nos modelos de regressão logística encontra-se no Apêndice B.

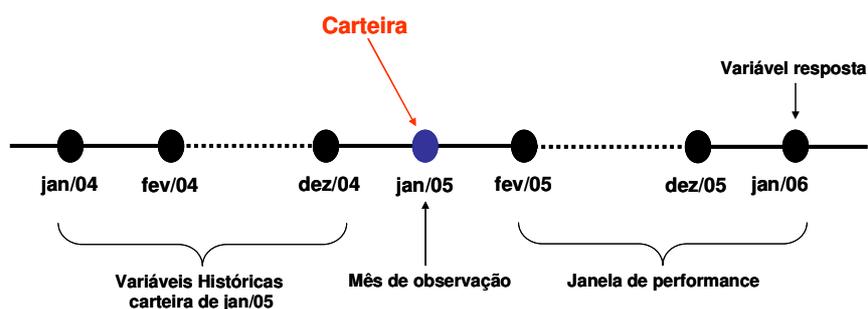


Figura 4.1 – Estrutura da base para os modelos de regressão logística

Para a estimação dos modelos de regressão logística foi utilizado o módulo Enterprise Miner, do software estatístico SAS³, amplamente utilizado no mercado. Uma observação importante é que os estimadores, nestes modelos, foram obtidos de forma a preverem os bons clientes, ou seja, quanto maior o escore gerado a partir dos modelos, melhor é o cliente com relação ao crédito ou à não tendência de cancelamento voluntário.

4.1.1 Modelos de probabilidade de default em 6 e 12 meses

Os modelos de default foram estimados de forma a tentar identificar as características de um bom cliente segundo as variáveis preditoras, sendo a resposta do modelo, portanto, a probabilidade de um cliente não apresentar problemas de default.

As variáveis selecionadas pelo método stepwise, assim como seus respectivos parâmetros e erros padrões, são mostradas na Tabela 4.1, para o modelo L1 e na Tabela 4.2, para o modelo L2. A variável *variante* não se mostrou significativa para ambos os modelos e foi excluída. A variável *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* não se mostrou significativa para o modelo com resposta em 6 meses.

³ <http://www.sas.com>

Tabela 4.1 – Estimadores para default em 6 meses (Modelo L1)

Variável	Estimador	Erro Padrão	P-Value
Intercepto	5.4127	0.1934	0.0000
Média exponencial de meses com status atrasado nos últimos 6 meses	-0.2173	0.0129	0.0000
Média exponencial da quantidade de compras nos últimos 6 meses	0.0909	0.0219	0.0000
Média exp.l do percentual de utilização do limite de crédito nos últimos 6 meses	-0.3191	0.0280	0.0000
Maior percentual de utilização do limite de crédito nos últimos 12 meses	-0.1386	0.0295	0.0000
Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0.0393	0.0159	0.0136

Tabela 4.2 – Estimadores para default em 12 meses (Modelo L2)

Variável	Estimador	Erro Padrão	P-Value
Intercepto	4.6852	0.1537	0.0000
Média exponencial de meses com status atrasado nos últimos 6 meses	-0.1884	0.0117	0.0015
Média exponencial da quantidade de compras nos últimos 6 meses	0.0849	0.0180	0.0000
Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	-0.3034	0.0235	0.0000
Maior percentual de utilização do limite de crédito nos últimos 12 meses	-0.1432	0.0236	0.0000
Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0.0413	0.0130	0.0000
Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0.0230	0.0100	0.0217

Observando-se os parâmetros dos modelos, as variáveis, em geral, parecem obedecer à lógica de crédito. As variáveis *média exponencial de meses com status atrasado nos últimos 6 meses*, *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* apresentaram parâmetros negativos em ambos os modelos, o que faz sentido, pois quanto maior for o valor destas variáveis, maior o grau de endividamento dos clientes e a tendência a apresentarem problemas de pagamento. A variável *média exponencial da quantidade de compras nos últimos 6 meses* apresentou parâmetro positivo, o que indica que uma maior quantidade de compras efetuadas contribui negativamente para a probabilidade de default, o que vai, de certa forma, contra a lógica das variáveis de utilização do limite de crédito. A variável *máxima quantidade consecutiva de meses sem compras nos últimos 12 meses* apresentou parâmetro negativo. A variável *quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* apresentou parâmetro negativo, no modelo L2, indicando que uma forte utilização do crédito rotativo indica uma menor tendência de o cliente não apresentar problemas de inadimplência.

4.1.2 Modelos de probabilidade de cancelamento em 6 e 12 meses

Os modelos de cancelamento foram estimados de forma a tentar identificar as características de um cliente que não apresentará problemas de cancelamento voluntário (attrition) segundo as variáveis preditoras, sendo a resposta do modelo, portanto, a probabilidade de um cliente não cancelar o cartão de crédito no horizonte de tempo observado.

As variáveis selecionadas pelo método stepwise, assim como seus respectivos parâmetros e erros padrões, são mostradas na Tabela 4.3, para o modelo L3 e na Tabela 4.4, para o modelo L4. As variáveis *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* e *maior percentual de utilização do limite de crédito nos últimos 6 meses* não se mostraram significativas nos modelos com resposta em 6 e 12 meses. A variável *média exponencial da quantidade de compras nos últimos 6 meses* não se mostrou significativa no modelo com resposta em 6 meses.

Tabela 4.3 – Estimadores para cancelamento em 6 meses (Modelo L3)

Variável	Estimador	Erro Padrão	P-Value
Intercepto	1.3625	0.1567	0.0000
Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0.0452	0.0128	0.0004
Média exponencial de meses com status atrasado nos últimos 6 meses	-0.2081	0.0137	0.0000
Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0.4766	0.0247	0.0000
Variante 1	-0.5456	0.1452	0.0002
Variante 2	0.0342	0.1254	0.7851
Variante 3	0,0000	.	.

Tabela 4.4 – Estimadores para cancelamento em 12 meses (Modelo L4)

Variável	Estimador	Erro Padrão	P-Value
Intercepto	1.2198	0.1484	0.0000
Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0.0762	0.0112	0.0000
Média exponencial de meses com status atrasado nos últimos 6 meses	-0.1499	0.0122	0.0000
Média exponencial da quantidade de compras nos últimos 6 meses	-0.0544	0.0215	0.0113
Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0.3659	0.0183	0.0000
Variante 1	-0.4927	0.1179	0.0000
Variante 2	-0.1136	0.0999	0.2555
Variante 3	0,0000	.	.

De uma forma geral, o resultado dos modelos de regressão logística para prever o cancelamento do relacionamento com a instituição por parte do cliente apresentou resultados coerentes com a lógica de relacionamento nos modelos de 6 (L3) e 12 meses (L4). O parâmetro negativo da variável *máxima quantidade consecutiva de meses sem compras nos últimos 12 meses* indica que quanto maior a quantidade de meses sem a utilização do produto, maior a probabilidade de o cliente cancelar o cartão de crédito. A variável *média exponencial de meses com status atrasado nos últimos 6 meses* também apresenta parâmetro negativo, o que pode ser interpretado como o fato de que um cliente que apresentou problemas de atraso recentemente tem uma maior tendência ao cancelamento do produto para evitar maiores problemas futuros devido a descontrole dos gastos. A variável *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses*, fortemente relacionada à utilização recente do produto, indica que quanto maior a utilização do limite de crédito, menor a probabilidade de o cliente cancelar o produto e o parâmetro positivo confirmou esta relação. Os parâmetros da variável *variante* em ambos os modelos, indicam que os clientes com a variante 1, aos quais são atribuídos os menores limites, possuem uma maior tendência ao cancelamento do produto em relação aos demais. A variável *média exponencial da quantidade de compras nos últimos 6 meses* apresentou resultado contraditório no modelo de 12 meses, pois se espera que quanto maior a quantidade de compras recente, menor a tendência de o cliente cancelar o produto, o que não é observado no sinal negativo do parâmetro.

4.2 Estimadores Modelo Multi-estado de Markov

Para a estimação do modelo multi-estado de Markov utilizou-se o pacote MSM, implementado em R⁴. O R é um software livre destinado à computação estatística e construído de forma colaborativa com muitos desenvolvedores. O pacote MSM, desenvolvido por Christopher Jackson (Jackson, 2006), permite estimar modelos multi-estado de Markov em tempo contínuo.

⁴ <http://cran.r-project.org/>

Os estimadores para cada tipo de transição, mostrados na Tabela 4.5, representam a relação entre as variáveis e o risco de transição entre os diversos estados. Os estados *em dia*, *rotativo*, *em atraso*, *cancelamento voluntário* e *default* são representados, respectivamente, pelos números 1, 2, 3, 4 e 5. Para calcular cada intensidade de transição entre os diversos estados utiliza-se (2.8), onde $q_{rs}^{(0)}$ é a função de base (*baseline*) para a transição *r-s*, e β_{rs}^T é o estimador associado à variável $x_i(t)$ para a transição *r-s*. A tabela com os *baselines* assim como as quantidades de todas as transições observadas na base de desenvolvimento se encontram no Apêndice C.

Tabela 4.5 – Estimadores modelo multi-estado de Markov

Transição	Variável	Estimativa	Erro Padrão	P-Value
1-2	Variante	-0,03047	0,03062	0,15982
1-2	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	0,14620	0,00503	0,00000
1-2	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0,02888	0,00718	0,00003
1-2	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0,11930	0,01240	0,00000
1-2	Média exponencial de meses com status atrasado nos últimos 6 meses	0,04471	0,00948	0,00000
1-2	Média exponencial da quantidade de compras nos últimos 6 meses	0,01012	0,00903	0,13124
1-2	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,10450	0,01091	0,00000
1-3	Variante	-0,01259	0,05174	0,40387
1-3	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	0,06376	0,01025	0,00000
1-3	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	0,00073	0,01048	0,47208
1-3	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0,04896	0,02139	0,01104
1-3	Média exponencial de meses com status atrasado nos últimos 6 meses	0,09908	0,01470	0,00000
1-3	Média exponencial da quantidade de compras nos últimos 6 meses	-0,00451	0,01599	0,38901
1-3	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,04720	0,01794	0,00425
1-4	Variante	-0,00170	0,06546	0,48962
1-4	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	0,03898	0,01451	0,00361
1-4	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	0,03621	0,01198	0,00125
1-4	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	-0,00358	0,02753	0,44832
1-4	Média exponencial de meses com status atrasado nos últimos 6 meses	-0,06060	0,02668	0,01155
1-4	Média exponencial da quantidade de compras nos últimos 6 meses	-0,08920	0,02327	0,00006
1-4	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,01480	0,02247	0,25503
2-1	Variante	-0,00408	0,03258	0,45017
2-1	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0,08829	0,00503	0,00000
2-1	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0,02233	0,01279	0,00107
2-1	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	-0,03854	0,01475	0,00450
2-1	Média exponencial de meses com status atrasado nos últimos 6 meses	-0,03142	0,00746	0,00001
2-1	Média exponencial da quantidade de compras nos últimos 6 meses	0,02735	0,00913	0,00137
2-1	Maior percentual de utilização do limite de crédito nos últimos 12 meses	-0,01829	0,01380	0,09245
2-3	Variante	-0,24240	0,03054	0,00000
2-3	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0,02538	0,00496	0,00000
2-3	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	0,00893	0,00660	0,08778
2-3	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0,02572	0,01401	0,03322
2-3	Média exponencial de meses com status atrasado nos últimos 6 meses	0,04792	0,00595	0,00000
2-3	Média exponencial da quantidade de compras nos últimos 6 meses	-0,02601	0,00960	0,00337
2-3	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,03647	0,01794	0,00411
2-4	Variante	-0,08053	0,08983	0,18501
2-4	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	0,03139	0,01425	0,01382
2-4	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0,06310	0,02185	0,00194
2-4	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	-0,02300	0,04060	0,28554
2-4	Média exponencial de meses com status atrasado nos últimos 6 meses	-0,06522	0,02306	0,00234
2-4	Média exponencial da quantidade de compras nos últimos 6 meses	-0,01988	0,02675	0,22867
2-4	Maior percentual de utilização do limite de crédito nos últimos 12 meses	-0,01635	0,03943	0,33919
3-1	Variante	-0,23820	0,06056	0,00004
3-1	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0,13830	0,01253	0,00000
3-1	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0,01462	0,01279	0,12642
3-1	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	-0,00245	0,02447	0,46017
3-1	Média exponencial de meses com status atrasado nos últimos 6 meses	-0,03284	0,01048	0,00087
3-1	Média exponencial da quantidade de compras nos últimos 6 meses	0,04227	0,02206	0,02767
3-1	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,02176	0,02427	0,18495
3-2	Variante	-0,01131	0,03679	0,37927
3-2	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0,00657	0,00686	0,16923
3-2	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0,03542	0,00840	0,00001
3-2	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0,02228	0,01575	0,07864
3-2	Média exponencial de meses com status atrasado nos últimos 6 meses	-0,04513	0,00670	0,00000
3-2	Média exponencial da quantidade de compras nos últimos 6 meses	0,04670	0,01315	0,00019
3-2	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,02129	0,01615	0,09368
3-4	Variante	-0,04925	0,08283	0,27606
3-4	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0,10300	0,02056	0,00000
3-4	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	0,05655	0,01793	0,00081
3-4	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	-0,10470	0,03603	0,00183
3-4	Média exponencial de meses com status atrasado nos últimos 6 meses	0,04588	0,01401	0,00053
3-4	Média exponencial da quantidade de compras nos últimos 6 meses	-0,04671	0,05006	0,17537
3-4	Maior percentual de utilização do limite de crédito nos últimos 12 meses	-0,09622	0,03432	0,00253
3-5	Variante	0,03015	0,04335	0,24339
3-5	Quantidade de meses com utilização de crédito rotativo nos últimos 12 meses	-0,01256	0,00823	0,06358
3-5	Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses	-0,03478	0,01003	0,00026
3-5	Média exp. do percentual de utilização do limite de crédito nos últimos 6 meses	0,04152	0,01899	0,01440
3-5	Média exponencial de meses com status atrasado nos últimos 6 meses	0,00357	0,00793	0,32602
3-5	Média exponencial da quantidade de compras nos últimos 6 meses	0,04935	0,01578	0,00088
3-5	Maior percentual de utilização do limite de crédito nos últimos 12 meses	0,03771	0,01959	0,02713

4.2.1 Transição *em dia* para *rotativo* (1-2)

Para as transições do tipo *em dia* para *rotativo* notamos, conforme Tabela 4.5, que as variáveis *variante* e *média exponencial da quantidade de compras nos últimos 6 meses* não são significativas. As variáveis *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses*, *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses*, *média exponencial de meses com status atrasado nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* mostraram-se significativas e com estimadores com sinal positivo, o que é esperado uma vez que estas variáveis indicam forte utilização do produto e dificuldade de pagamento, ou seja, quanto maior o valor, maior a probabilidade de migração para o estado *rotativo*. A variável *quantidade consecutiva de meses sem compras nos últimos 12 meses* mostrou-se significativa e com o estimador com o sinal negativo, o que indica que quanto maior a quantidade de meses consecutivos sem a utilização do produto, menor a propensão do cliente utilizar o crédito rotativo.

4.2.2 Transição *em dia* para *em atraso* (1-3)

No caso das transições do tipo *em dia* para *em atraso* a Tabela 4.5 mostra que as variáveis *variante*, *quantidade consecutiva de meses sem compras nos últimos 12 meses* e *média exponencial da quantidade de compras nos últimos 6 meses* não são significativas. As variáveis *quantidade de meses com utilização do crédito rotativo nos últimos 12 meses*, *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses*, *média exponencial de meses com status atrasado nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* são significativas e os estimadores apresentaram sinal positivo, o que é esperado pois valores maiores destas variáveis indicam maiores chances de o cliente apresentar problemas de atraso.

4.2.3 Transição *em dia* para *cancelamento voluntário* (1-4)

Para as transições do tipo *em dia* para *cancelamento voluntário* vemos que as variáveis *variante*, *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* se apresentaram não significativas, conforme a Tabela 4.5. A variável *quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* se mostrou significativa e com estimador positivo, o que indica que quanto maior o valor da variável, maior a propensão ao cancelamento voluntário. Isto pode dever-se ao fato de que o cliente vem apresentando dificuldades de pagamento do total da fatura e tende a cancelar o produto para não apresentar problemas de atraso. A variável *quantidade consecutiva de meses sem compras nos últimos 12 meses* mostrou-se significativa e com o estimador positivo, o que indica que períodos longos sem utilização do produto aumentam as chances de cancelamento voluntário. As variáveis *média exponencial de meses com status atrasado nos últimos 6 meses* e *média exponencial da quantidade de compras nos últimos 6 meses* apresentaram estimadores negativos e significativos, o que indica que quanto maior o valor destas variáveis, menor a propensão ao cancelamento do produto por parte do cliente. A variável de atraso nos últimos meses indica que o cliente tem grande propensão a apresentar problemas de atraso futuramente e, estando em atraso, maior propensão a apresentar problemas de default como podemos ver na Tabela 4.5, ou seja, o cliente tende a ir para o estado absorvente de default e não para o estado absorvente de cancelamento voluntário. A variável de quantidade de compras nos últimos meses indica que quanto maior a quantidade de compras nos meses recentes, menor o risco de cancelamento voluntário do produto, o que é esperado, pois mostra que o cartão de crédito está sendo bastante utilizado.

4.2.4 Transição *rotativo* para *em dia* (2-1)

A variável *variante* mostrou-se não significativa na estimativa da transição do tipo *rotativo* para *em dia*. A variável *média exponencial da quantidade de compras nos últimos 6 meses* se mostrou significativa e com o estimador positivo, o que indica que quanto maior a quantidade de compras nos meses recentes, maior a probabilidade de o cliente retornar para o estado em dia. Este resultado não é

intuitivo uma vez que quanto maior a quantidade de compras, mais endividado o cliente fica e, conseqüentemente, com maior dificuldade para pagar o valor total da fatura. As variáveis *quantidade de meses com utilização de crédito rotativo nos últimos 12 meses*, *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses* e *média exponencial de meses com status atrasado nos últimos 6 meses* se mostraram significativas e com estimadores negativos, ou seja, quanto maiores os valores das variáveis, maiores as dificuldades de os clientes migrarem do estado *rotativo* para *em dia*, resultados esperados uma vez que estas variáveis indicam o grau de endividamento dos clientes. A variável *quantidade consecutiva de meses sem compras nos últimos 12 meses* apresentou estimador negativo, resultado não esperado, pois quanto maior a quantidade de meses sem compras menor o grau de endividamento, o que deveria intuitivamente levar o cliente de volta ao estado *em dia*.

4.2.5 Transição *rotativo* para *em atraso* (2-3)

Na análise do efeito das variáveis nas transições do tipo *rotativo* para *em atraso* todas as variáveis mostraram-se significativas conforme podemos ver na Tabela 4.5. A variável *variante*, que em todos os outros tipos de transições não se mostrou significativa, exceto nas transições do tipo *em atraso* para *em dia*, obteve um estimador significativo e negativo, com o módulo do estimador muito superior ao módulo das demais variáveis. Quanto maior a variante, ou seja, quanto maior o nível de renda, menor o risco de o cliente apresentar problemas de atraso dado que está utilizando o crédito rotativo, fato esperado uma vez que clientes com maiores níveis de renda têm maiores possibilidades de suportar o financiamento de suas dívidas através do recurso de pagamento mínimo. A variável *quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* apresentou estimador com o sinal negativo, indicando que o fato de o cliente possuir histórico de utilização do crédito rotativo diminui as chances de migração para o estado *em atraso*. As variáveis *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses*, *média exponencial de meses com status atrasado nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* têm estimadores com o sinal positivo, resultado esperado, pois valores elevados destas

variáveis indicam alto grau de endividamento podendo levar o cliente a não capacidade de pagamento de suas dívidas e conseqüentes problemas de atrasos. A variável *média exponencial da quantidade de compras nos últimos 6 meses* obteve estimador negativo, resultado de difícil interpretação, podendo hipoteticamente indicar que o cliente valoriza o produto como forma de pagamento e financiamento de dívidas efetuando várias compras de pequeno valor, mas intuitivamente espera-se que uma maior quantidade de compras levaria mais facilmente o cliente a apresentar problemas de atraso.

4.2.6 Transição *rotativo* para *cancelamento voluntário* (2-4)

No caso das transições do tipo *rotativo* para cancelamento voluntário, evento raro na base observada (Apêndice C), observa-se que apenas as variáveis *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses*, *quantidade consecutiva de meses sem compras nos últimos 12 meses* e *média exponencial de meses com status atrasado nos últimos 6 meses* se mostraram significativas, conforme a Tabela 4.5.

4.2.7 Transição *em atraso* para *em dia* (3-1)

Para as transições do tipo *em atraso* para *em dia*, que indicam a recuperação financeira do cliente, a variável *variante* apresentou estimador significativo e distante de zero. O fato de valores maiores de variantes estarem associados a um menor poder de recuperação pode indicar que clientes de nível de renda mais elevados, ao apresentarem problemas de atrasos, têm maiores dificuldades de recuperação. As variáveis *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* e *média exponencial de meses com status atrasado nos últimos 6 meses* apresentaram estimadores com sinais negativos, indicando que clientes com maior histórico de utilização do crédito rotativo ou atrasos possuem menor capacidade de recuperação. A variável *média exponencial da quantidade de compras nos últimos 6 meses* apresentou estimador positivo. As demais variáveis não se mostraram significativas, de acordo com a Tabela 4.5.

4.2.8 Transição em atraso para rotativo (3-2)

Para as transições do tipo *em atraso* para *rotativo*, que de certa forma também podem indicar um grau de recuperação do cliente, verifica-se que as variáveis *variante* e *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* não são significativas, conforme vemos na Tabela 4.5. As variáveis *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses*, *média exponencial da quantidade de compras nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* apresentaram estimadores significativos e com sinais positivos, indicando que valores maiores destas variáveis estão associados a menores chances de transição para o estado *rotativo*, dado que se está no estado *em atraso*. A variável *média exponencial de meses com status atrasado nos últimos 6 meses* apresentou estimador significativo e com sinal negativo, indicando uma não tendência de recuperação do cliente conforme aumenta-se o valor da variável. A variável *quantidade consecutiva de meses sem compras nos últimos 12 meses* apresentou estimador significativo e negativo.

4.2.9 Transição em atraso para cancelamento voluntário (3-4)

Nas transições do tipo em atraso para cancelamento voluntário, as variáveis *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses*, *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* apresentaram estimadores significativos e negativos enquanto que as variáveis *quantidade consecutiva de meses sem compras nos últimos 12 meses* e *média exponencial de meses com status atrasado nos últimos 6 meses* apresentaram estimadores significativos e positivos, de acordo com a Tabela 4.5. As demais variáveis não se mostraram significativas.

4.2.10 Transição *em atraso para default* (3-5)

Para as transições do tipo *em atraso para default*, as variáveis *variante e média exponencial de meses com status atrasado nos últimos 6 meses* não se mostraram significativas, conforme Tabela 4.5. As variáveis *máxima quantidade de meses com utilização de crédito rotativo nos últimos 12 meses* e *quantidade consecutiva de meses sem compras nos últimos 12 meses* apresentaram estimadores significativos e com sinais negativos. As variáveis *média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses*, *média exponencial da quantidade de compras nos últimos 6 meses* e *maior percentual de utilização do limite de crédito nos últimos 12 meses* apresentaram estimadores significativos e com sinais positivos, o que indica que estas variáveis, quando em níveis elevados, indicam maior risco de os clientes migrarem para o estado absorvente *default*.

4.2.11 Comentários gerais

De uma forma geral, vemos que todas as variáveis se mostraram significativas para pelo menos dois tipos de transições, indicando que poderíamos ter perda de informações no caso de eliminarmos da análise algumas destas variáveis. A variável *variante*, por exemplo, se apresentou bastante significativa para dois tipos de transições e não significativa para os oito outros tipos de transições.

4.3 Comparação entre o Modelo Multi-estado de Markov e a Regressão Logística

O principal objetivo deste trabalho é testar a aplicação do modelo multi-estado de Markov em cartões de crédito, um produto com características de eventos recorrentes multi-estado. Para confrontar o modelo multi-estado de Markov com outras alternativas de análise, foram gerados 4 modelos de regressão logística com resposta binária, utilizando as mesmas informações e bases de dados de desenvolvimento e validação disponíveis para o modelo multi-estado.

Uma característica importante do modelo multi-estado de Markov é o fato de que uma vez estimada a matriz de intensidades de transição podemos facilmente gerar vários modelos de score, para diversos horizontes de tempo, sendo possível ordenar os clientes de acordo com seus perfis de risco para diversos interesses. No presente estudo em cartões de crédito temos condições de gerar modelos de behaviour scoring, anti-attrition scoring, collection scoring e modelos de propensão à utilização de crédito rotativo. Para comparação com a regressão logística foram gerados modelos de probabilidade de default em 6 e 12 meses, que podem ser considerados modelos de behaviour scoring uma vez que variáveis comportamentais no produto são utilizadas e modelos de probabilidade de cancelamento voluntário do produto em 6 e 12 meses, que podemos considerar como modelos de anti-attrition scoring.

Para gerar os modelos de default, a partir da matriz de intensidades de transição, foram utilizadas as probabilidades de transição de qualquer estado não absorvente para o evento absorvente default, durante um período de 6 meses e 12 meses. No caso dos modelos de cancelamento utilizando a matriz de intensidades de transição foram utilizadas as probabilidades de transição, também de qualquer estado não absorvente para o evento cancelamento voluntário do produto, durante um período de 6 meses e 12 meses.

Além dos modelos de default e cancelamento foram analisados também modelos de atraso no pagamento de faturas e modelos de propensão à utilização do

crédito rotativo ao longo de períodos de 6 e 12 meses. No caso dos modelos de atraso utilizou-se a probabilidade de transição durante os períodos de 6 e 12 meses, de qualquer estado não absorvente, para o evento *em atraso*. Para os modelos de propensão à utilização do crédito rotativo foi utilizada a probabilidade de transição também durante os períodos de 6 e 12 meses, de qualquer estado não absorvente, para o evento utilização de crédito rotativo.

4.3.1 Indicadores de Desempenho

A Tabela 4.6 mostra as estatísticas de Kolmogorov-Smirnov e os coeficiente de Gini para cada um dos modelos analisados. As Figuras 4.2 a 4.9 ilustram a capacidade de ordenação dos clientes através de gráficos de *back test*. Todos os indicadores de performance foram obtidos na amostra de validação.

Tabela 4.6 – Indicadores de desempenho dos modelos

Escore	Estatística de Kolmogorov-Smirnov	Coefficiente de Gini
Escore MSM Default 6 meses	55,7%	70,0%
Escore LOG Default 6 meses	51,9%	64,0%
Escore MSM Default 12 meses	50,4%	57,6%
Escore LOG Default 12 meses	46,7%	54,5%
Escore MSM Cancelamento 6 meses	47,9%	52,8%
Escore LOG Cancelamento 6 meses	54,1%	59,3%
Escore MSM Cancelamento 12 meses	37,0%	41,6%
Escore LOG Cancelamento 12 meses	41,0%	45,2%
Escore MSM Atraso 6 meses	42,8%	39,3%
Escore MSM Atraso 12 meses	38,7%	31,1%
Escore MSM Rotativo 6 meses	49,3%	33,1%
Escore MSM Rotativo 12 meses	45,1%	26,9%

Os modelos de probabilidade de transição para default utilizando a matriz de transição do modelo multi-estado, para 6 e 12 meses, apresentaram resultados superiores aos obtidos pelos modelos de probabilidade de default através da regressão logística, de acordo com os indicadores da Tabela 4.6. No caso de probabilidade de cancelamento os modelos obtidos através da regressão logística apresentaram resultados superiores, também mostrado na Tabela 4.6.

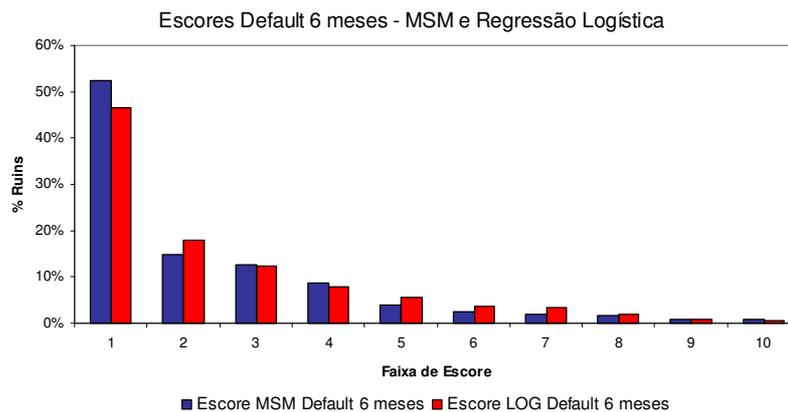


Figura 4.2 – Gráfico de back test dos modelos de escore de default em 6 meses

Na Figura 4.2 podemos verificar que os modelos de default em 6 meses apresentam uma ordenação consistente, com uma pequena vantagem para o modelo obtido via matriz de transição. Observamos que nas classes de menores escores temos maiores percentuais de clientes ruins e que este percentual cai consistentemente conforme se avança nas faixas maiores de escore. A estatística de Kolmogorov Smirnov e o índice de Gini também indicam vantagem para o modelo multi-estado de Markov.

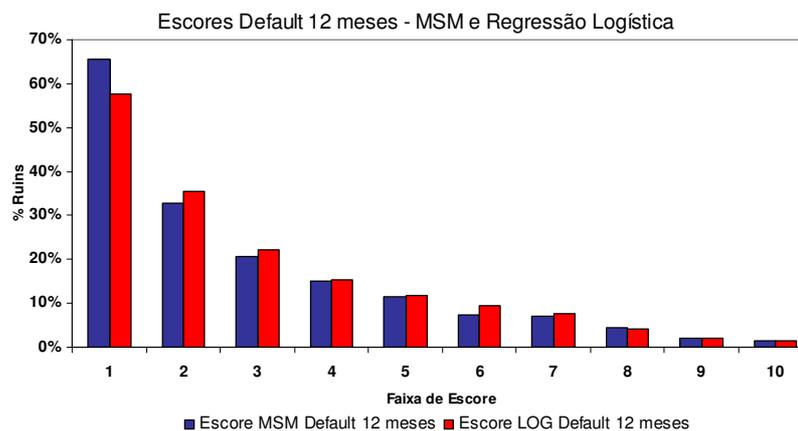


Figura 4.3 – Gráfico de back test dos modelos de escore de default em 12 meses

Os modelos de default em 12 meses, apesar de um desempenho um pouco inferior aos modelos de 6 meses, também apresentaram bons resultados, novamente com pequena vantagem para o modelo multi-estado de Markov, conforme podemos ver na Tabela 4.6 e nos gráficos da Figuras 4.3.

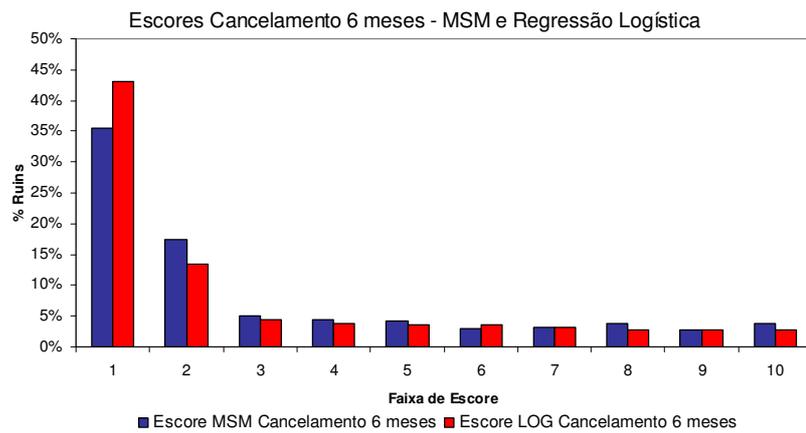


Figura 4.4 – Gráfico de back test dos modelos de escore de cancelamento em 6 meses

Nos modelos de cancelamento em 6 meses observamos resultados satisfatórios nos dois tipos de modelos, com vantagem para o obtido via regressão logística. O gráfico de back test da Figura 4.4 mostra que os modelos apresentam uma melhor ordenação nos menores escores enquanto que nos maiores escores a queda no percentual de cancelamento não ocorre de forma acelerada. No caso dos modelos de 12 meses ainda observamos traçados satisfatórios da queda no percentual de cancelamento conforme melhora-se a faixa de escore, como podemos observar na Figura 4.5, apesar de as estatísticas de Kolmogorov Smirnov e os índices de Gini terem apresentado resultados bastante inferiores aos verificados nos modelos de 6 meses.

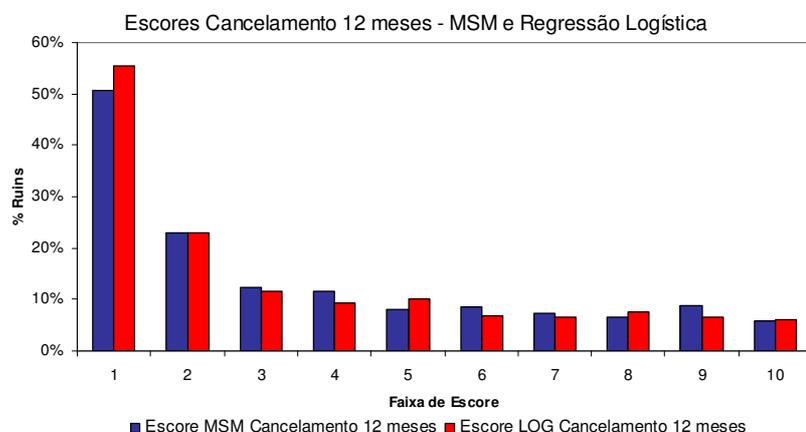


Figura 4.5 – Gráfico de back test dos modelos de escore de cancelamento em 12 meses

O modelo de atraso em 6 meses, obtido a partir da matriz de intensidades de transição do modelo multi-estado de Markov apresentou uma ordenação consistente, conforme vemos na Figura 4.6, onde os clientes com os menores escores de fato apresentaram proporcionalmente mais problemas de atraso no produto no horizonte de tempo de 6 meses.

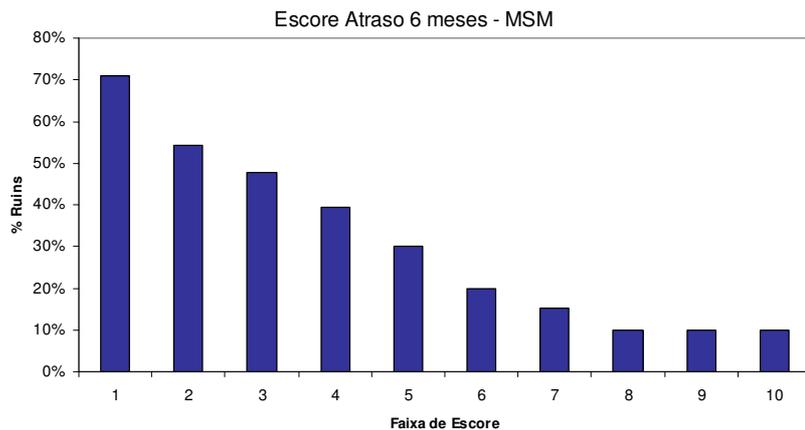


Figura 4.6 – Gráfico de back test do modelo de escore de atraso em 6 meses

No caso do modelo de atraso em 12 meses, obtido a partir da matriz de intensidades de transição do modelo multi-estado de Markov, ainda temos algum grau de ordenação dos clientes segundo seus escores de atraso, apesar de nas 3 melhores faixas de escore não observarmos uma ordenação consistente, conforme mostrado na Figura 4.7.

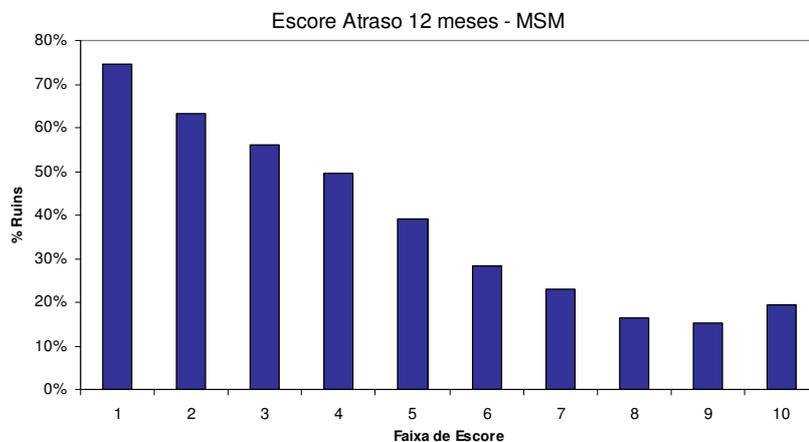


Figura 4.7 – Gráfico de back test do modelo de escore de atraso em 12 meses

O modelo de utilização de crédito rotativo em 6 meses, obtido a partir da matriz de intensidades de transição do modelo multi-estado de Markov, apresenta algum grau de ordenação, conforme mostrado na Figura 4.8, apesar da inversão observada na faixa dos 10% piores escores em relação à faixa seguinte. É possível tomar ações de incentivo à utilização do produto ou retenção do cliente utilizando este modelo, onde quanto maior o escore, menor a propensão em relação à utilização do crédito rotativo.

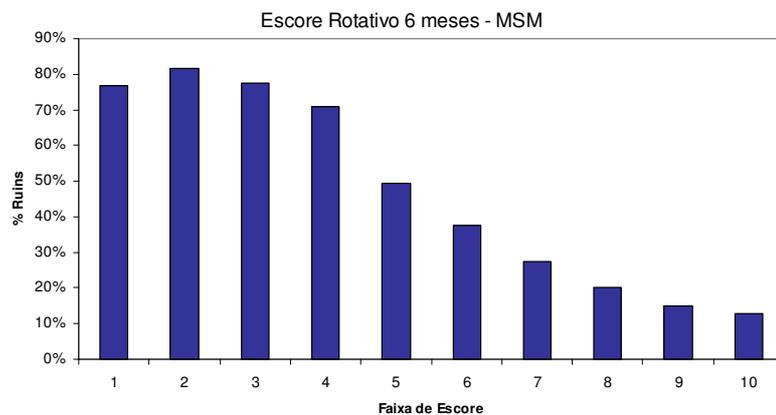


Figura 4.8 – Gráfico de back test do modelo de escore de utilização do crédito rotativo em 6 meses

No caso do modelo de utilização do crédito rotativo em 12 meses, o gráfico de back test mostrado na Figura 4.9 indica um modelo com algum poder de discriminação, porém com um grau de diferenciação dos clientes não muito forte em faixas de escore próximas.

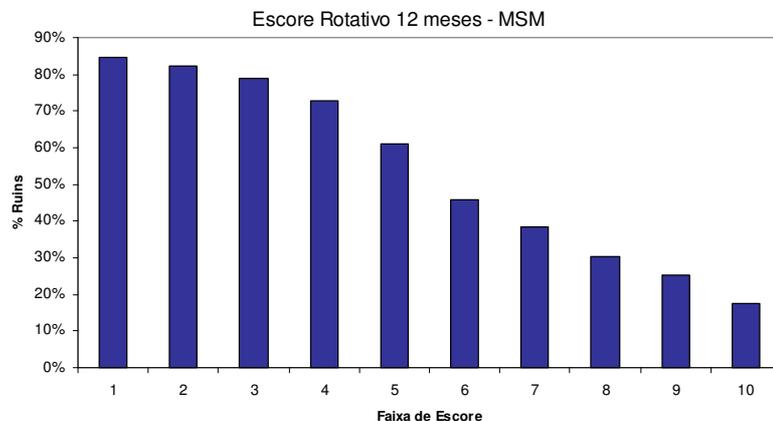


Figura 4.9 – Gráfico de back test do modelo de escore de utilização do crédito rotativo em 12 meses

Capítulo 5 - Conclusão

Neste trabalho foi estudada a aplicação do modelo multi-estado de Markov em tempo contínuo em cartões de crédito, aproveitando as características de eventos multi-estado e recorrência do uso do produto ao longo do relacionamento com a instituição financeira. Foi verificado o desempenho de alguns dos possíveis modelos de escore obtidos a partir da utilização da matriz de intensidades de transição, tais como modelos de escore de default, modelos de escore de cancelamento, modelos de escore de atraso e modelos de escore de utilização do crédito rotativo.

Para confrontar com os modelos de escore de default e cancelamento, utilizando a mesma base de dados e a mesma seleção de variáveis, foram estimados modelos de regressão logística múltipla com resposta binária e verificou-se vantagem de desempenho para os modelos obtidos a partir da matriz de intensidades de transição no caso de escores de default e vantagem para os modelos obtidos a partir da regressão logística para os escores de cancelamento.

Uma característica bastante interessante do modelo multi-estado de Markov é o fato de que uma vez estimada a matriz de intensidades de transição podemos facilmente gerar vários modelos de escore, para diversos horizontes de tempo, sendo possível ordenar os clientes de acordo com seus perfis de risco para diversos interesses. Esse tipo de modelo pode ser testado em qualquer produto que possua as características de eventos recorrentes multi-estado.

Melhores modelos de behaviour ou anti-attrition scoring utilizando tanto a regressão logística quanto o modelo multi-estado de Markov poderiam ser desenvolvidos, aproveitando-se melhor as características comportamentais dos clientes, assim como analisando diversas outras variáveis como, por exemplo, informações comportamentais do cliente no mercado, que poderiam ser obtidas junto a bureaus de mercado (ACSP e Serasa, por exemplo). A aplicação de técnicas de seleção de variáveis mais apropriadas, tanto para a regressão logística quanto

para o modelo multi-estado de Markov, também poderia proporcionar modelos de melhor performance, dado um conjunto de muitas variáveis explicativas.

O estudo do efeito de diversas variáveis nas transições entre ratings de empresas, fornecidos pelas agências de classificação, ou nas transições entre ratings internos de empresas fica como sugestão para estudos futuros, assim como a aplicação de metodologias mais apropriadas para a seleção de variáveis.

Apêndice A

Lista de Variáveis Analisadas

Tabela A.1 – Lista de Variáveis Analisadas

Variável
ID
Status (<i>em dia, rotativo, em atraso, cancelamento voluntário, default</i>)
Tempo desde o início de observação das transições
Bandeira
Cobertura (nacional / internacional)
Limite de Crédito
Maior percentual de utilização do limite de crédito nos últimos 12 meses
Maior quantidade de compras nos últimos 12 meses
Máxima quantidade consecutiva de meses com status <i>em dia</i> nos últimos 12 meses
Máxima quantidade consecutiva de meses com status <i>rotativo</i> nos últimos 12 meses
Máxima quantidade consecutiva de meses com status <i>rotativo</i> ou <i>em atraso</i> nos últimos 12 meses
Máxima quantidade consecutiva de meses sem compras nos últimos 12 meses
Média da quantidade de compras nos últimos 12 meses
Média do percentual de utilização do limite de crédito nos últimos 6 meses
Média exponencial da quantidade de compras nos últimos 6 meses
Média exponencial de meses com status atrasado nos últimos 6 meses
Média exponencial de meses com status em dia nos últimos 6 meses
Média exponencial de meses com status rotativo nos últimos 6 meses
Média exponencial do percentual de utilização do limite de crédito nos últimos 6 meses
Meses desde a ativação
Quantidade de meses com status <i>em atraso</i> nos últimos 12 meses
Quantidade de meses com status <i>em dia</i> nos últimos 12 meses
Quantidade de meses com utilização do crédito rotativo nos últimos 12 meses
Quantidade de transições do tipo <i>em atraso</i> para <i>em dia</i> nos últimos 12 meses
Quantidade de transições do tipo <i>em atraso</i> para <i>rotativo</i> nos últimos 12 meses
Quantidade de transições do tipo <i>em dia</i> para <i>em atraso</i> nos últimos 12 meses
Quantidade de transições do tipo <i>em dia</i> para <i>rotativo</i> nos últimos 12 meses
Quantidade de transições do tipo <i>em dia</i> para <i>rotativo</i> ou <i>em atraso</i> nos últimos 12 meses
Quantidade de transições do tipo <i>rotativo</i> para <i>em atraso</i> nos últimos 12 meses
Quantidade de transições do tipo rotativo para em dia nos últimos 12 meses
Variante

A Tabela A.1 mostra a lista das variáveis analisadas neste trabalho, sendo que 5 são variáveis cadastrais e 23 são variáveis comportamentais no produto

Apêndice B

Exemplos de Bases de Dados

Modelo Multi-estado de Markov

Na Figura B.1 é ilustrado um trecho da base de desenvolvimento do modelo multi-estado de Markov. Podemos reparar que um mesmo indivíduo aparece em diversos tempos na base, podendo ter suas variáveis variando ao longo do tempo.

id	status	tempo	variante	status_rev	consec_compzero	explim	expatr
26	1	0	2	6	0	7	1
26	2	1	2	5	0	7	1
26	1	2	2	6	0	7	1
26	2	3	2	6	0	6	1
26	2	4	2	7	0	6	1
26	2	5	2	7	0	6	1
26	2	6	2	8	0	8	1
26	1	7	2	8	1	7	1
26	2	8	2	8	1	6	1
26	2	9	2	9	1	6	1
26	2	10	2	9	1	5	1
26	2	11	2	9	1	6	1
26	2	12	2	9	1	8	1
27	2	0	2	10	1	9	1
27	2	1	2	11	1	9	1
27	2	2	2	11	1	9	1
27	2	3	2	12	1	9	1
27	2	4	2	12	1	9	1
27	2	5	2	12	1	9	1
27	2	6	2	12	1	8	1
27	2	7	2	12	1	8	1
27	2	8	2	12	1	8	1
27	3	9	2	12	1	8	1
27	5	10	2	11	1	8	8
28	1	0	2	0	12	1	1
28	4	1	2	0	12	1	1
29	1	0	2	0	12	3	1
29	1	1	2	0	12	3	1

Figura B.1 – Trecho da base de dados utilizada no desenvolvimento do modelo multi-estado de Markov

Modelo Regressão Logística

A Figura B.2 mostra um trecho da base de dados utilizada para o desenvolvimento de um dos modelos de regressão logística. Podemos reparar que cada indivíduo e suas respectivas variáveis aparecem em uma única linha na base de dados.

id	performance	variante	status_rev	consec_compzero	explim	expatr	expcomp
63	3	2	8	12	4	9	1
64	6	2	3	11	4	1	1
65	6	2	1	0	6	1	4
66	6	3	0	7	3	1	3
67	6	2	3	8	6	1	7
68	6	2	1	1	6	1	4
69	6	2	3	0	4	6	4
70	3	1	2	2	7	9	2
71	6	2	0	4	5	1	2
72	3	1	1	6	5	1	2
73	6	2	4	2	8	1	4
74	6	1	10	1	9	4	3
75	6	2	0	1	6	1	3
76	6	2	0	8	1	1	1
77	6	1	3	3	6	1	3
78	3	1	6	8	10	7	2
79	6	2	10	3	10	1	7
80	6	2	1	1	3	1	5
81	6	2	3	1	4	1	3
82	6	2	8	3	4	1	2
83	6	2	0	7	4	1	2
84	6	3	0	0	5	1	7
85	6	2	0	0	8	1	8
86	6	2	0	1	6	1	7
87	6	2	3	12	1	1	1
88	6	1	2	8	3	6	2
89	6	1	0	1	6	1	2
90	6	1	0	1	4	1	3

Figura B.2 – Trecho da base de dados utilizada no desenvolvimento de um dos modelos de regressão logística

Apêndice C

Baselines e Transições

Baselines

A Tabela C.1 mostra os *baselines* para cada tipo de transição.

Tabela C.1 – *Baselines* para as transições

Transição	Estimativa	Erro Padrão	P-Value
1-2	0,01447	0,00146	0,00000
1-3	0,01108	0,00057	0,00000
1-4	0,01596	0,00355	0,00000
2-1	0,35420	0,03954	0,00000
2-3	0,16300	0,01809	0,00000
2-4	0,03127	0,01074	0,00180
3-1	0,26450	0,05546	0,00000
3-2	0,26480	0,03582	0,00000
3-4	0,14940	0,05086	0,00166
3-5	0,11030	0,01813	0,00000

Transições

A Figura C.1 mostra a quantidade observada de cada tipo de transição, ao longo de 12 meses, entre os diversos estados possíveis na base de desenvolvimento do modelo multi-estado de Markov (1 - *em dia sem utilização de crédito rotativo*, 2 - *em dia com utilização de crédito rotativo*, 3 - *em atraso*, 4 - *cancelamento voluntário* e 5 - *default*). As quantidades da Figura C.2 não refletem as quantidades reais de transições da instituição da qual foi retirada.

	to				
from	1	2	3	4	5
1	57662	4617	1760	912	0
2	4580	18616	3675	157	0
3	1100	2808	3111	340	1680

Figura C.1 – Quantidade de transições entre os diversos estados possíveis na base de desenvolvimento do modelo multi-estado de Markov

Apêndice D – Programa MSM- R

Neste apêndice, apresentamos comandos do programa MSM utilizados no processo de estimação do Modelo Multi-Estado de Markov.

No trecho abaixo estão os comandos que geram a matriz de intensidades de transição e estimam o modelo multi-estado de Markov.

```
> library("msm")
> base3_msm<-read.table(file="c:\\msm\\fase3\\txt\\base3_msm.txt",header=TRUE)
> statetable.msm(status, id, data=base1_msm)
> inits.q<-rbind(c(0.25, 0.25, 0.25, 0.25, 0),c(0.25, 0.25, 0.25, 0.25, 0),c(0.2, 0.2, 0.2, 0.2, 0.2), c(0, 0, 0, 1, 0),c(0, 0, 0, 0, 1))
> inicial.q<-crudeinits.msm(status~tempo, id, data=base3_msm, qmatrix=inits.q)
> base3.msm<-msm(status~tempo, subject=id, data=base3_msm, qmatrix=inicial.q, death=c(4,5),
exacttimes=TRUE, covariates=~variante+status_rev+consec_compzero+explim+expatr+expcomp+maxlim)
```

Com a matriz de intensidades de transição, foi criada a função *escora_msm_r*, em R, para calcular as probabilidades de transição em 6 e em 12 meses, para cada indivíduo da amostra de validação:

```
`escora_msm_r` <-
function(base)
{
  n <- nrow(base)
  mat<-matrix(nrow=n, ncol=34)
  for(i in 1:n)
  {
    p6m=pmatrix.msm(base3.msm, t=6, covariates=list(variante=base[i,2],
status_rev=base[i,3], consec_compzero=base[i,4], explim=base[i,7], expatr=base[i,8], expcomp=base[i,9],
maxlim=base[i,10]))
    p12m=pmatrix.msm(base3.msm, t=12, covariates=list(variante=base[i,2],
status_rev=base[i,3], consec_compzero=base[i,4], explim=base[i,7], expatr=base[i,8], expcomp=base[i,9],
maxlim=base[i,10]))

    mat[i,1]<-base[i,1]
    mat[i,2]<-base[i,5]
    mat[i,3]<-base[i,11]
    mat[i,4]<-base[i,12]

    mat[i,5]<-p6m[1,1]
    mat[i,6]<-p6m[1,2]
    mat[i,7]<-p6m[1,3]
    mat[i,8]<-p6m[1,4]
    mat[i,9]<-p6m[1,5]
    mat[i,10]<-p6m[2,1]
    mat[i,11]<-p6m[2,2]
    mat[i,12]<-p6m[2,3]
    mat[i,13]<-p6m[2,4]
    mat[i,14]<-p6m[2,5]
```

```

mat[i,15]<-p6m[3,1]
mat[i,16]<-p6m[3,2]
mat[i,17]<-p6m[3,3]
mat[i,18]<-p6m[3,4]
mat[i,19]<-p6m[3,5]

mat[i,20]<-p12m[1,1]
mat[i,21]<-p12m[1,2]
mat[i,22]<-p12m[1,3]
mat[i,23]<-p12m[1,4]
mat[i,24]<-p12m[1,5]
mat[i,25]<-p12m[2,1]
mat[i,26]<-p12m[2,2]
mat[i,27]<-p12m[2,3]
mat[i,28]<-p12m[2,4]
mat[i,29]<-p12m[2,5]
mat[i,30]<-p12m[3,1]
mat[i,31]<-p12m[3,2]
mat[i,32]<-p12m[3,3]
mat[i,33]<-p12m[3,4]
mat[i,34]<-p12m[3,5]

}
return(mat)
}

valida<-escora_msm_r(basevalida);

```

Utilizando a base de validação foi gerada uma sintaxe em SAS para transformar as probabilidades em um escore, que varia de 0 a 1000, onde quanto maior o escore, menor a probabilidade de transição. As variáveis respostas para o cálculo dos indicadores de desempenho dos modelos também são calculadas nesse programa. Após o cálculo dos escores para o modelo multi-estado de Markov, são inseridos na base os escores calculados pelos modelos de regressão logística, em uma escala de 0 a 1000. A base gerada possui todos os resultados dos modelos de escore utilizados nesta dissertação.

```

/* Calcula escores */

data msm.base3_escores_geral;
merge msm.base3_escorada_tzero_r (in = a drop=status_6 status_12)
      msm.base3_msm_valida_tzero_status (in = b keep=id status_1-status_12);
by id;
if a and b;

/* escore enquadrado 6 e 12 meses */
if status = 1 then escore_msm_enq_6m=1000-round(p6m15*1000,1);
  if status = 2 then escore_msm_enq_6m=1000-round(p6m25*1000,1);
  if status = 3 then escore_msm_enq_6m=1000-round(p6m35*1000,1);

if status = 1 then escore_msm_enq_12m=1000-round(p12m15*1000,1);
  if status = 2 then escore_msm_enq_12m=1000-round(p12m25*1000,1);
  if status = 3 then escore_msm_enq_12m=1000-round(p12m35*1000,1);

/* escore cancelado 6 e 12 meses */
if status = 1 then escore_msm_can_6m=1000-round(p6m14*1000,1);
  if status = 2 then escore_msm_can_6m=1000-round(p6m24*1000,1);
  if status = 3 then escore_msm_can_6m=1000-round(p6m34*1000,1);

if status = 1 then escore_msm_can_12m=1000-round(p12m14*1000,1);
  if status = 2 then escore_msm_can_12m=1000-round(p12m24*1000,1);

```

```

        if status = 3 then escore_msm_can_12m=1000-round(p12m34*1000,1);

/* escore atrasado 6 e 12 meses */
if status = 1 then escore_msm_atr_6m=1000-round(p6m13*1000,1);
    if status = 2 then escore_msm_atr_6m=1000-round(p6m23*1000,1);
    if status = 3 then escore_msm_atr_6m=1000-round(p6m33*1000,1);

if status = 1 then escore_msm_atr_12m=1000-round(p12m13*1000,1);
    if status = 2 then escore_msm_atr_12m=1000-round(p12m23*1000,1);
    if status = 3 then escore_msm_atr_12m=1000-round(p12m33*1000,1);

/* escore revolving 6 e 12 meses */
if status = 1 then escore_msm_rev_6m=1000-round(p6m13*1000,1);
    if status = 2 then escore_msm_rev_6m=1000-round(p6m23*1000,1);
    if status = 3 then escore_msm_rev_6m=1000-round(p6m33*1000,1);

if status = 1 then escore_msm_rev_12m=1000-round(p12m13*1000,1);
    if status = 2 then escore_msm_rev_12m=1000-round(p12m23*1000,1);
    if status = 3 then escore_msm_rev_12m=1000-round(p12m33*1000,1);

enq6m = 0;
if status_6 = 5 then enq6m = 1;
enq12m = 0;
if status_12 = 5 then enq12m = 1;

can6m = 0;
if status_6 = 4 then can6m = 1;
can12m = 0;
if status_12 = 4 then can12m = 1;

array status_vet{*} status_1 - status_12;

rev6m=0; atr6m=0; rev12m=0; atr12m=0;
i=1;
do while (i<=12);

    if i<=6 then do;
        if status_vet{i} = 2 then rev6m = 1;
        if status_vet{i} = 3 then atr6m = 1;
    end;

    if status_vet{i} = 2 then rev12m = 1;
    if status_vet{i} = 3 then atr12m = 1;

    i=i+1;

end;

drop i;
run;

/* resultados regressões logísticas */

/* escores enquadramento */
data msm.logistic_enq6m;
    set emdata.Sd_3osgb;
    escore_log_enq_6m=round(P_perfenq6m6*1000, 1);
    keep id escore_log_enq_6m;
run;
data msm.logistic_enq12m;
    set emdata.Sd_656j2;
    escore_log_enq_12m=round(P_perfenq12m6*1000, 1);
    keep id escore_log_enq_12m;
run;

/* escores cancelamento */
data msm.logistic_can6m;
    set emdata.Sd_78vpt;
    escore_log_can_6m=round(P_perfcan6m6*1000, 1);
    keep id escore_log_can_6m;
run;
data msm.logistic_can12m;
    set emdata.Sd_fw85i;
    escore_log_can_12m=round(P_perfcan12m6*1000, 1);
    keep id escore_log_can_12m;

```

```
run;

/* Insere escores regressões logísticas na base de escores geral */

data msm.base3_escores_geral;
  merge msm.base3_escores_geral (in = a)
        msm.logistic_enq6m (in = b)
        msm.logistic_enq12m (in = c)
        msm.logistic_can6m (in = d)
        msm.logistic_can12m (in = e);
  by id;
  if a and b and c and d and e;
run;
```

Referências Bibliográficas

ABECS (2007). Associação Brasileira das Empresas de Cartões de Crédito e Serviços. Disponível em: <http://www.abecs.org.br>. Acesso em: 07.06.2007.

Blackwell, M. and Sykes, C. (1992). The Assignment of Credit Limits with a Behaviour-scoring System. *IMA Journal of Mathematics Applied in Business and Industry*, **12**. 293-310.

Bussab, W. O. e Morettin, P. A. (2005). *Estatística Básica*. 5 ed. Saraiva: São Paulo.

Canal Executivo (2007). *Crédito ao consumidor deve chegar a R\$ 310 bi em 10 anos*. Disponível em: <http://www2.uol.com.br/canalexecutivo/notas05/130920054.htm>. Acesso em: 23.12.2006.

Clipping Ministério do Planejamento (2006). *Por R\$ 1 bi, Bradesco compra American Express no Brasil*. Disponível em: <http://clipping.planejamento.gov.br/Noticias.asp?NOTCod=256446>. Acesso em: 26.12.2006.

Coelho, D.C. (2006). *Um Modelo de Previsão para a Renda Utilizando Esquemas de Censura Complexos*. Dissertação de Mestrado. Universidade de São Paulo.

Colosimo, E. A. e Giolo, S. R. (2006). *Análise de Sobrevivência Aplicada*. 1 ed. Edgard Blücher Ltda: São Paulo.

Commenges, D. (1999). *Multi-state Models in Epidemiology*. Université de Bordeaux, France.

Cook, R. J. and Lawless, J. F. (2002). *Analysis of Repeated Events*. University of Waterloo, Canada.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. Chapman and Hall: London.

Figueiredo, R. P. (2006). *A Evolução do Sistema de Pagamentos Brasileiro e o Desaparecimento do Cheque: Realidade ou Exagero?*, Dissertação de Mestrado. Ibmec São Paulo.

Folha Online (2004). *Unibanco vende participação na Credicard para Itaú e Citibank*. Disponível em: <http://www1.folha.uol.com.br/folha/dinheiro/ult91u90502.shtml>. Acesso em 23.12.2006.

Folha Online (2006). *Marca Credicard ficará só com Citibank a partir de 2009*. Disponível em: <http://www1.folha.uol.com.br/folha/dinheiro/ult91u112957.shtml>. Acesso em 23.12.2006.

Hoper, M. A. and Lewis, E. M. (1992). Behaviour Scoring and Adaptive Control Systems. In *Credit Scoring and Credit Control*, ed L. C. Thomas, J. N. Crook, D. B. Edelman, Claredons Press: Oxford.

Hougaard, P. (1999). *Multi-state Models: A Review*. Novo Nordisk: Denmark.

Hosmer, D.W. e Lemeshow, S. (2000). *Applied Logistic Regression*. 2ed. John Wiley & Sons: New York.

Jackson, C. H. (2006). *Multi-state modelling with R: The MSM Package Version 0.6*. Imperial College: London.

Jackson, C. H. and Sharples, L. D. (2002). Hidden Markov Models for the Onset and Progression of Bronchiolitis Obliterans Syndrome in Lung Transplant Recipients. *Statistics in Medicine*, **21**. 113-128.

Jackson, C.H, Sharpless, L.D., Thompson, S.G., Duffy, S.W. e Couto, E. (2003). Multistate Markov Models for Disease Progression with Classification Error. *The Statistician*, **52**. 193-209.

Jonhson, R. A. and Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*. 6 ed. Prentice-Hall: New York.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The Analysis of Panel Data Under a Markov Assumption. *Journal of the American Statistical Association*, **80**. 863-871.

Kay, J. (1986). A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies. *Biometrics*, **42**. 855-865.

Lewis, E. M. (1992). *An Introduction to CreditScoring*. 2 ed. FairIsaac and Co., Inc.

Marshall, G. and Jones, R. H. (1995). Multi-state Markov Models and Diabetic Retinopathy. *Statistics in Medicine*, **14**. 1975-1983.

Neto, F. L. (2004). *A Hybrid Scale Intensity Model for Recurrent Event Data*. Universidade Federal de São Carlos: São Carlos.

Oliveira, W. (2000). *CRM & E-Business*. Visual Business: Florianópolis.

Ohtoshi, C. (2003). *Uma Comparação de Regressão Logística, Árvores de Classificação e Redes Neurais: Analisando Dados de Crédito*. Dissertação de Mestrado. Universidade de São Paulo.

Paes, A. T. (1999). *Modelos Semiparamétricos Para Eventos Recorrentes*. Dissertação de Mestrado. Universidade de São Paulo.

Paula, G. A. (2004). *Modelos de Regressão com Apoio Computacional, Versão Preliminar*. Universidade de São Paulo: São Paulo.

Pereira, G. H. A. (2004). *Modelos de Risco de Crédito de Clientes: Uma Aplicação a Dados Reais*. Dissertação de Mestrado. Universidade de São Paulo.

Portal Economia (2006). Disponível em <http://www.portaleconomia.com.br>. Acessado em: 23.12.2006.

Quidim, I. L. (2005). *Análise de Sobrevivência com Fração de Fidelizados: Uma Aplicação na Área de Marketing*. Dissertação de Mestrado. Universidade de São Paulo.

Rosa, P. T. M. (2000). *Modelos de Credit Scoring Regressão Logística Chaid e Real*. Dissertação de Mestrado. Universidade de São Paulo.

Servigny, A. and Renault, O. (2004). *Measuring and Managing Credit Risk*. McGraw-Hill: New York.

Sinhorigno, S. (2007). *Previsão de Inadimplência de Transações com Cartão de Crédito*, Dissertação de Mestrado. Universidade de São Paulo.

Thomas, L. C. Edelman, D. B. and Crook, J. N. (2002). *Credit Scoring and Its Applications*. Siam: Philadelphia.

Thomas, L. C., Ho, J. and Scherer, W. T. (2001). Time Will Tell: Behaviour Scoring and the Dynamics of Consumer Credit Assessment. *IMA Journal of Management Mathematics*. **12**. 89-103.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)